Essays on Macroeconomic Analyses with Factor Models

Taiki Yamamura

A thesis submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

June 2019

Statement of originality

I, Taiki Yamamura, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: 19/June/2019

Details of collaboration and publications

Chapter 3 is joint work with Professor Emmanuel Guerre and I contributed 90% of this work.

Abstract

The thesis contains three essays, which are related to macroeconomic analyses with factor models.

In Chapter 1, I investigate time variations in the monetary policy effects on the economy in Japan, by using time-varying Factor Augmented Vector Autoregression (FAVAR) model. The main interest is whether and how the policy effects change due to the Bubble Burst and during the (near-)zero interest rate period. As an analysis methodology, I propose and adopt the following two-stage procedure. In the first stage, the shadow rate is estimated by a non-linear term structure model, where the shadow rate represents a policy stance of the monetary policy authority during the (near-)zero interest rate period. Using the estimated rate as a policy instrument, the second stage estimates the time-varying FAVAR model.

Chapter 2 investigates the performance of time-varying FAVAR in terms of whether it correctly captures time variations in monetary policy transmission to macro-economy. The analysis is conducted through Monte Carlo (MC)-based experiments, and the model's performance is examined in comparison with that of time-varying VAR which does not use unobserved factors. The experiments show that the time-varying FAVAR adequately detects the time variations even under a situation where the time-varying VAR fails to do this. Using the computation techniques proposed in the recent literature, the above result is interpreted in terms of the information sufficiency of those two empirical models.

In Chapter 3, my attention moves to an estimation of factor model with machine-learning approach. The chapter is devoted to proposing a novel method to identify a grouped factor structure by introducing a l_1-constraint (Lasso approach) of the pair-wise difference of the factor loadings. Through theoretical analyses including Monte Carlo experiments, the advantage of the Lasso-based method is revealed over the existing methods.

Acknowledgment

I would like to express my gratitude to my supervisors, Haroon Mumtaz, Renato Faccini and Emmanuel Guerre, for their invaluable supervision and continuous support. Their immense motivation and patience have provided me a power and spirit to conduct the academic study at Queen Mary.

Besides my supervisors, I would like to give my appreciation to my colleagues at Queen Mary for enjoying and sharing our stimulating research life with me. Special thanks should go to Mirela Miescu, Germano Ruisi, and Shayan Zakipour-Saber. I would also like to thank Michele Piffer for providing several attractive research works and sharing a lot of stimulating discussions.

Last, but not least, I would like to dedicate this thesis to my beloved wife and daughter, Michiko and Reona. To Michiko, thank you for marrying me and giving us a beautiful baby. To my parents and parents-in-law, I never forget everything of your endless patience and support.

Contents

List of Figures

List of Tables

Chapter 1

Changes in Monetary Policy Effects through the Bubble Burst and the Zero Interest Rate Regime in Japan

1.1 Introduction

During these few decades, many countries have experienced several structural changes in their economy. One well-known example is the Great Moderation, which is a marked decline in the volatility of macroeconomic variables such as output and inflation. As reported by many researchers, the Great Moderation occurs during the 1980s and the subsequent periods in the U.S. (see, for example, Perez-Quiros and McConnell (2000) and Blanchard and Simon (2001)). In UK, a similar phenomenon is observed after the introduction of inflation targeting in 1992[1]. Many research studies also claim that monetary policy effects on the economy significantly change through this phenomenon. Famous works include Lubik and Schorfheide (2004) and Boivin and Giannoni (2006), whose studies are based on sticky-price Dynamic Stochastic General Equilibrium (DSGE). In the more recent literature, similar implications are provided by Baumeister et al. (2013) and Ellis et al. (2014) who use time-varying

[1]See, for example, Ellis et al. (2014).

Factor Augmented Vector Autoregression (FAVAR) framework[2]. As discussed in those papers, it is arguably the case that the major structural changes in the economy significantly affect transmission of monetary policy effects throughout the economy.

As an extension of the works by Baumeister et al. (2013) and Ellis et al. (2014), I apply the time-varying FAVAR model to Japan's data from 1975 to 2007. Whereas those earlier works focus on the Great Moderation as a source of changing monetary policy effects, the interest of this study is whether and how the policy effects change due to the Bubble Burst and the (near-)zero interest rate policy[3].

In the literature, it is common to use short-term interest rate as monetary policy instrument. However, it is not necessarily relevant in this study, because the rate is close to the lower bound during the (near-)zero interest rate period. To deal with this issue, I conduct a two-step analysis. In the first step, I evaluate *shadow rate* from 1995 to 2007 in Japan, where the shadow rate represents policy stance of the monetary policy authorities during the (near-)zero interest rate regime[4]. Then, using the shadow rate as policy instrument, the second step is devoted to estimating the time-varying FAVAR model, by which I examine how monetary policy effects change from the 1970s to the 2000s.

In the first step, I use a non-linear term structure model called SRTSM (shadow rate term structure model). This model is originally proposed by Black (1995). The model introduces the shadow rate s_t that is linear in Gaussian factors, with the actual short-term interest rate r_t defined by the maximum of the shadow rate and zero (i.e. $r_t = \max(s_t, 0)$). By using a state-space framework of latent factors, the literature has been analytically studying relation between the shadow rate and the yields on assets of various maturities. For a while, the literature finds the analytical solution only in the case of one factor framework, which does not necessarily match empirical data in a good manner. However, Wu and Xia (2016) succeed in deriving the solution for a multi-factor

[2]They analyze U.S. and UK data from the 1970s to the 2000s. Each study finds the evidence of a significant shift in monetary policy effects due to the Great Moderation.

[3]Japan experienced the Bubble Burst around 1990-91, by which both of the housing and financial markets plunged. To tackle the adverse impact of this crisis on the economy, the Bank of Japan kept on lowering the policy rate. The rate was lowered to 0.5 % in 1995, since which it has been near zero (i.e. near-zero interest rate policy).

[4]Note that the data period ends at 2007, and it is due to availability of the yield curve data of the government bond. For detail, see Section 1.2.

framework, by which an empirical application of SRTSM becomes much more realistic.

As for the second step, it should be mentioned that time-varying VAR (Vector Autoregression) can be also a relevant analysis model. However, as pointed out by Benati and Surico (2009) and Graeve (2017), the time-varying VAR is not necessarily suitable for identifying time variations in monetary policy transmission. In this perspective, Yamamura (2018) also claims that the time-varying FAVAR behaves in an adequate manner even under the situation that the time-varying VAR does not[5]. Based on those works, I choose to use the time-varying FAVAR throughout this study.

The main findings from the estimated FAVAR are the following. The impulse response of the real output to monetary policy shock does not change significantly across the whole study period. However, the response of the aggregate price becomes stronger during the early 1990s, and the timing of this change is associated with the Bubble Burst. Regarding the policy effects during the (near-)zero interest rate period, no significant evidence of their time variation is observed. A simulation work with the New Keynesian model implies that the above observation can be explained by shifts in the underlying dynamics of the macro-economy such as a transition in the monetary policy rule. Using the estimated FAVAR, I also examine monetary policy propagation to disaggregate prices, and it reveals that dispersion of the price responses rapidly increases during the early 1990s.

This chapter is organized as follows. In Section 1.2, the shadow rate is estimated. Section 1.3 introduces the time-varying FAVAR model and its estimation methodology. Section 1.4 shows the analysis results, and Section 1.5 provides a conclusion.

1.2 Estimation of shadow rate

This section is devoted to estimating the shadow rate during the (near-)zero interest rate regime in Japan.

[5] Yamamura (2018)'s work is reported in Chapter 2 of this thesis.

1.2.1 Shadow rate model

During the (near-)zero interest rate regime, the Bank of Japan often exercised an unconventional policy such as large-scale asset purchasing which is expected to affect the long-term interest rates. It is not straightforward to asses the impact of such a policy, because the short-term interest rate is always near its lower bound during the zero interest rate period. In the literature on the term structure models, many studies have tried to deal with this issue. One popular approach is the Gaussian affine term structure model (GATSM). It describes the relationship between the yields on assets with different maturities, by using the (unobserved) Gaussian factors. However, the model describes the nominal interest rates using a linear function of the latent factors, by which they are allowed to become negative. This is a severe problem when the short-term interest rate is close to the zero lower boundary.

To tackle this drawback the shadow rate term structure model (SRTSM) is proposed by Black (1995). In this model, the short-term interest rate r_t is defined by the maximum of *the shadow rate* s_t and the lower bound r_{min} (i.e. $r_t = \max(s_t, r_{min})$). For a while, the literature finds the analytical solution only in the case of one factor framework, which does not necessarily match empirical data in a good manner. However, Wu and Xia (2016) succeed in deriving the solution for a multi-factor framework, by which an empirical application of SRTSM becomes much more realistic.

1.2.2 Equation system

In the SRTSM, the short term interest rate r_t is defined by the maximum of the shadow rate s_t and the lower bound r_{min}:

$$r_t = \max(s_t, r_{min}) \tag{1.1}$$

where the shadow rate s_t is defined by an affine function of the unobserved factors X_t:

$$s_t = \delta_0 + \delta_1' X_t \tag{1.2}$$

The latent factors X_t are assumed to follow a VAR(1) process under the physical measure $\mathbb{P}$:

$$X_{t+1} = \mu + \rho X_t + \Sigma\varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, I) \tag{1.3}$$

When $y_{n,t}$ denotes the zero-coupon yield at time t for a loan maturing at $t+n$, the no-arbitrage assumption makes an asset price being a martingale under the risk-neutral measure $\mathbb{Q}$:

$$y_{n,t} = -\frac{1}{n}\log(E_t^Q[\exp(-\int_t^{t+n} r_\tau d\tau)])$$

where $E_t^Q[\cdot]$ is the conditional expectation under the $\mathbb{Q}$-measure. Under this measure, the factors X_t are written as

$$X_{t+1} = \mu^Q + \rho^Q X_t + \Sigma^Q \varepsilon_{t+1}^Q, \qquad \varepsilon_{t+1} \overset{Q}{\sim} N(0, I) \tag{1.4}$$

Note that the two measures $\mathbb{P}$ and $\mathbb{Q}$ are related by

$$\begin{aligned} \mu - \mu^Q &= \Sigma\lambda_0 \\ \rho - \rho^Q &= \Sigma\lambda_1 \end{aligned}$$

where the parameters λ_0 and λ_1 are related to the market price of risk[6].

Based on Eqs. (1.1)-(1.4), Wu and Xia (2016) derive an analytical approximation of forward rate as follows:

$$\begin{aligned} f_{n,n+1,t} &= r_{min} + \sigma_n^Q g\left(\frac{a_n + b_n' X_t - r_{min}}{\sigma_n^Q}\right) \\ (\sigma_n^Q)^2 &= Var_t^Q(s_{t+n}) = \sum_{j=0}^{n-1} \delta_1'(\rho^Q)^j \Sigma\Sigma'(\rho^{Q'})^j \delta_1 \end{aligned}$$

where $f_{n,n+1,t}$ is the forward rate at time t for a loan starting at $t+n$ and maturing at $t+n+1$. The function $g(\cdot)$ is $g(z) = z\Phi(z) + \phi(z)$ where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function (CDF) and probability density function (PDF) of a standard normal distribution, respectively. The definitions of a_n and b_n are

$$\begin{aligned} \bar{a}_n &= \delta_0 + \delta_1'(\sum_{j=0}^{n-1}(\rho^Q)^j)\mu^Q \\ a_n &= \bar{a}_n - \frac{1}{2}\delta_1'(\sum_{j=0}^{n-1}(\rho^Q)^j)\Sigma\Sigma'(\sum_{j=0}^{n-1}(\rho^Q)^j)'\delta_1 \\ b_n &= \delta_1'(\rho^Q)^n \end{aligned}$$

Using all the above, the equation system of the shadow rate model can

[6]Following Duffee (2002), the market price of risk λ_t is assumed to be essentially affine in terms of X_t:

$$\lambda_t = \lambda_0 + \lambda_1 X_t$$

be summarized by the following state space representation:

$$f_{n,n+1,t} = r_{min} + \sigma_n^Q g\left(\frac{a_n + b_n' X_t - r_{min}}{\sigma_n^Q}\right) \tag{1.5}$$

$$\begin{aligned} X_{t+1} &= \mu + \rho X_t + \Sigma\varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0,I) \\ X_{t+1} &= \mu^Q + \rho^Q X_t + \Sigma^Q \varepsilon_{t+1}^Q, \qquad \varepsilon_{t+1} \overset{Q}{\sim} N(0,I) \end{aligned} \tag{1.6}$$

1.2.3 Identification scheme

In the literature, it is widely accepted that three factors can sufficiently account for the cross-sectional variation in the zero-coupon yields[7]. On this basis, three factor models are used throughout this exercise ($\dim(X_t) = 3$). For identification of the model, the following restrictions are imposed[8]:

(i) $\delta_1 = [1, 1, 0]'$

(ii) $\mu^Q = 0$

(iii) Σ is lower triangular.

(iv) ρ^Q is in real Jordan form with the eigenvalues in the descending order. When assuming that ρ^Q has three distinct eigenvalues, the two smaller eigenvalues are almost identical to one another (see Creal and Wu (2015) and Wu and Xia (2016)). Furthermore, ρ_1^Q is assumed to be one in this exercise ($\rho_1^Q = 1$)[9]. From the above assumptions, ρ^Q obtains the following representation:

$$\rho^Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho_2^Q & 1 \\ 0 & 0 & \rho_2^Q \end{pmatrix}$$

1.2.4 Estimation methodology

Estimation procedure The SRTSM is estimated by the Bayesian approach. The estimation algorithm consists of the following steps. To

[7]The three factors are expected to capture intercept, slope and curvature of the yield curve, respectively.

[8]These restrictions are used by many works in the literature (see, for example, Joslin et al. (2011), Hamilton and Wu (2014), and Wu and Xia (2016)).

[9]The relevancy of this assumption is discussed in appendix 1.6.1.

begin with, given the initial values for the factors, ρ_2^Q, Σ and δ_0 are sampled by using the scheme of Jacquier et al. (1994), and R is drawn from the inverse gamma distribution. The factors X_t are sampled using the Kalman filter. It should be mentioned that the observation equation (1.5) is not linear in X_t. To cope with this non-linearity, the *extended* Kalman filter is used[10]. Finally, using Eq. (1.6), the remaining parameters μ and ρ are estimated by the least square approach. The above procedure is repeated 10,000 times, where the first 9,000 iterations are discarded as a burn-in ((M,M_0)=(10000, 9000)).

Data set Wright (2011) provides the dataset of the one-month forward rates $f_{n,n+1,t}$[11], where the maturities n are: 1 and 3 months; and 1, 2, 5, 7, 10 and 12.5 years ($n = 1, 3, 12, 24, 60, 84, 120, 150$). The data period spans from January 1992 to December 2007 (1992:01-2007:12).

1.2.5 Estimated result

The estimated shadow rate is displayed in Fig. 1.1. In the figure, the actual short-term interest rate (overnight call rate) is also indicated by the black line. Before 1995, the shadow rate is almost equal to the call rate. However, in 2001-2006 (the period when the quantitative easing was conducted by the Bank of Japan), the shadow rate fluctuates in the negative region, showing a clear deviation from the call rate. Note also that the shadow rate becomes positive around March 2006, when the quantitative easing was terminated. As for the estimates of the model parameters, see Table 1.1 and Fig. 1.2. Furthermore, Fig. 1.3 shows the estimated shadow rates in several cases of iteration times (M, M_0), which confirm that the results converge sufficiently.

[10] For detail of the algorithm, see Wu and Xia (2016).

[11] In the paper, he estimates the zero-coupon yield curve for ten countries including Japan.

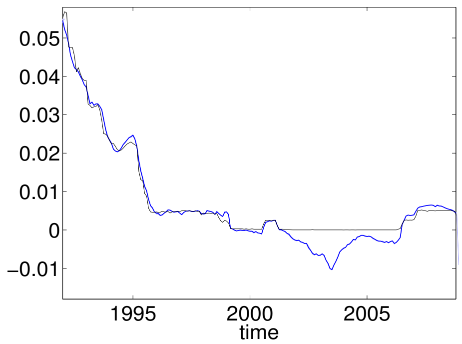


Figure 1.1: Estimated shadow rate: The blue line represents the estimated shadow rate. The black line is the overnight call rate.

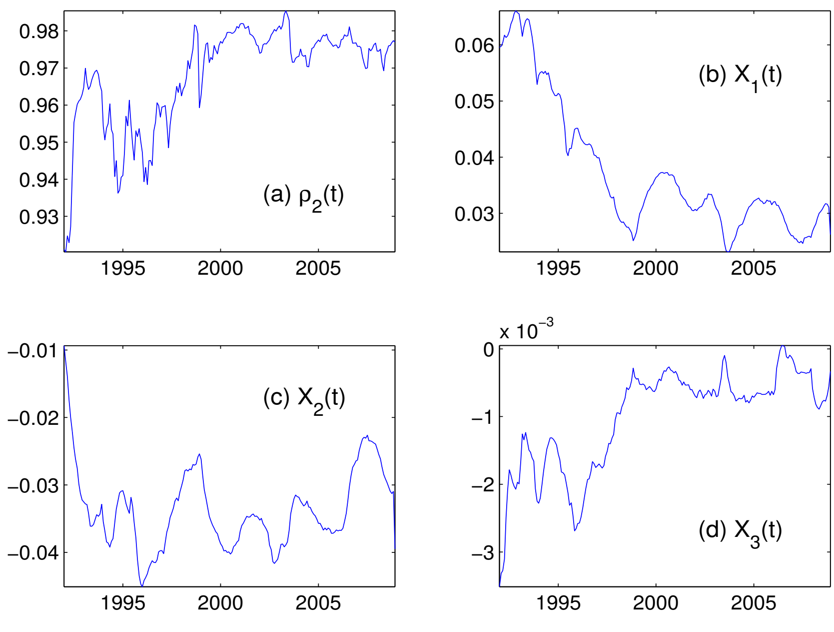


Figure 1.2: Estimated result for $\rho_2^Q(t)$ and X_t

Table 1.1: Estimated result for the model parameters in SRTSM

Parameters	Estimates
δ_0	$(4.26{\pm}0.03){\times}10^{-3}$
R	$(1.16{\pm}0.06){\times}10^{-7}$
Σ	$\begin{pmatrix} 2.06 \pm 0.01 & 0 & 0 \\ -2.66 \pm 0.01 & 2.25 \pm 0.01 & 0 \\ -0.02 \pm 0.00 & -0.15 \pm 0.00(1) & 0.44 \pm 0.00(2) \end{pmatrix} \times 10^{-4}$
μ	$\begin{pmatrix} 1.31 \pm 0.02 \\ -1.66 \pm 0.04 \\ -0.02 \pm 0.01 \end{pmatrix} \times 10^{-3}$
ρ	$\begin{pmatrix} 9.84 \pm 0.00(3) & 0.30 \pm 0.00(4) & 1.25 \pm 0.08 \\ 0.01 \pm 0.01 & 9.41 \pm 0.01 & -4.87 \pm 0.17 \\ 0.01 \pm 0.00(1) & -0.01 \pm 0.00(1) & 9.57 \pm 0.01 \end{pmatrix} \times 10^{-1}$

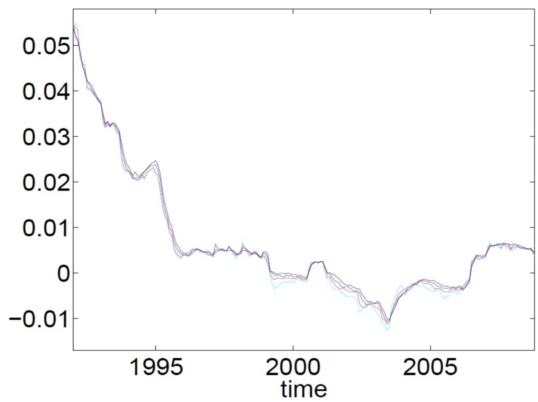


Figure 1.3: Stability check of the estimated shadow rate: The black, blue, red and cyan lines are the estimates with $(M, M_0) = (10000, 9000)$, $(8000, 7000)$, $(5000, 4500)$ and $(3000, 2700)$, respectively.

1.3 Time-varying FAVAR model with stochastic volatility

1.3.1 The model

The specification follows Baumeister et al. (2013) and Ellis et al. (2014). The model can be written in state space form. First, the observation equation is

$$X_t = \Lambda^f F_t + \Psi^R R_t + e_t \ , \quad e_t \sim N(0, R) \qquad (t = 1, ..., T) \tag{1.7}$$

where $X_t = [X_{1t} \ ... \ X_{Nt}]'$ is a panel of N observed variables, $F_t = [F_t^1 \ ... \ F_t^K]'$ denote K latent factors, and R_t is monetary policy instrument[12]. The disturbances $e_t = [e_{1t} \ ... \ e_{Nt}]'$ are *i.i.d.* with $E[e_t] = 0$ and $E[e_t e_t'] = R$, where the $N \times N$ matrix R is assumed to be diagonal. For unique identification of latent factors, it is assumed that the upper $K \times K$ block of Λ^f is identity matrix, and also that the upper $K \times 1$ block of Ψ^R is zero[13]. The transition equation is given by

$$Z_t = c_t + B_{1,t} Z_{t-1} + ... + B_{L,t} Z_{t-L} + v_t \ , \quad v_t \sim N(0, \Omega_t) \tag{1.8}$$

where Z_t denotes the common factors made up of latent factors F_t and interest rate R_t ($Z_t = [F_t' \ R_t]'$). Note that this equation allows for time-varying parameters (c_t, $B_{k,t}$ with $k = 1, ..., L$) and stochastic volatility (Ω_t). As a common choice in the literature, the lag length is set equal to two ($L = 2$)[14]. The volatility matrix Ω_t is factored as

$$\Omega_t = A_t^{-1} H_t (A_t^{-1})' \tag{1.9}$$

[12] As described in 1.3.3, the policy instrument R_t is defined by the overall call rate for the pre-1995 period (1975-1994), and by *the shadow rate* for the post-1995 period (1995-2007).

[13] For detail, see Bernanke et al. (2005).

[14] This choice is mainly motivated from a computational perspective. For more detail, see Primiceri (2005), Baumeister et al. (2013) and Ellis et al. (2014).

where H_t is diagonal and A_t is lower-triangular:

$$H_t = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t} \end{bmatrix} \quad , \quad A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1,t} & \cdots & a_{nn-1,t} & 1 \end{bmatrix} \tag{1.10}$$

Following Primiceri (2005), Baumeister et al. (2013) and Ellis et al. (2014), the time evolution of the VAR coefficients and volatility matrix is specified as

$$\begin{aligned} \beta_t &= \beta_{t-1} + \eta_t \\ \alpha_t &= \alpha_{t-1} + \tau_t \\ \log \sigma_t &= \log \sigma_{t-1} + \varepsilon_t \end{aligned} \tag{1.11}$$

where β_t stacks all of the VAR coefficients, α_t stacks non-zero and non-one elements of the matrix A_t, and σ_t stacks diagonal components of the matrix H_t. All of the innovations in the model are assumed to be jointly normally distributed, and their variance covariance matrices are specified by

$$V \equiv Var\left(\begin{bmatrix} u_t \\ \eta_t \\ \tau_t \\ \varepsilon_t \end{bmatrix}\right) = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & D & 0 \\ 0 & 0 & 0 & G \end{bmatrix} \tag{1.12}$$

where G is assumed to be diagonal ($G = \text{diag}(g_i)$).

1.3.2 Identification of monetary policy shock

As well as model specification, the identification scheme follows a standard recommendation in the literature (Bernanke et al. (2005), Baumeister et al. (2013) etc.). To identify a monetary policy shock, two restrictions are imposed into the model. First, the recursive identification scheme is applied to the transition equation. Note that the monetary policy instrument R_t is placed last in Z_t (see the discussion around Eq. (1.8)). With this ordering, monetary policy shock is assumed to be the only one structural shock which does not affect the other factors contemporaneously. Second, the macroeconomic series are divided into two categories: *slow-moving* and *fast-moving*. The slow-moving variable is a pre-determined variable in the current period, such as IIP and CPI. As for such variable, Ψ is assumed to be zero ($\Psi^{ii} = 0$), which is the second restriction to identify a monetary policy shock. On the other hand, the fast-moving variable is sensitive to contemporaneous news and shocks, such as asset prices, and as for such variables, Ψ is not assumed to be zero ($\Psi^{ii} \neq 0$). The categorization of slow/fast variables in our dataset will be given in 1.3.3.

A potential problem associated with the recursive identification is that estimation can be possibly biased due to incompatibility of zero restrictions with the underlying data-generating process (Canova and Pina (2005)). However, contemporaneous effect of monetary shock on the observed variables can be controlled by the factor loadings. Hence, the recursive identification scheme adopted here is less likely to impose a severe restriction on dynamics between the monetary policy instrument and the observed series.

1.3.3 Estimation methodology

Estimation of time-varying FAVAR The model is estimated by the Bayesian approach as described by Kim and Nelson (1999). The priors and starting values for the parameters follow a standard recommendation in the literature (Baumeister et al. (2013) etc.). The estimation algorithm is made up of the following steps. To begin with, the factor loadings $\Gamma \left(= (\Psi^R, \Lambda^f)\right)$ and the variance R $(= \text{diag}(R_i))$ are sampled conditionally on latent factors. Γ and R_i are drawn from the normal and inverse gamma distributions respectively. The VAR coefficients (β_t), the off-diagonal elements of A_t ($a_{ij,t}$), and the covariance matrices of disturbances in their random-walk process (Q and D) are subsequently drawn through the method developed by Carter and Kohn (1994), while $h_{i,t}$ and g_i are simulated using the scheme described in Jacquier et al. (1994). The latent factors F_t are sampled by the algorithm of Carter and Kohn (1994). The above steps are iterated 30,000 times, with the first 27,000 draws removed as a burn-in (M =30,000 and M_0=27,000). For more details of the algorithm, see appendices 1.6.2 and 1.6.3.

Computation of IRF (impulse response function) The impulse responses of the factors Z_t (F_t and R_t) to a monetary policy shock are computed at each point in time in the data sample. To make the responses comparable over time, the shock is normalized so as that the contemporaneous response of the monetary policy instrument R_t is 100 basis points (i.e. 1%) at each point in time. Following Koop et al. (1996), the responses of the factors Z_t at time t for horizon h are defined as:

$$(IRF)^Z_{t,h} = E[Z_{t+h}|\Xi_t, Z^{t-1}, \mu_{MP}] - E[Z_{t+h}|\Xi_t, Z^{t-1}]$$

where $\Xi_{i,t}$ represents all the parameters and hyperparameters of the VAR at time t, Z^{t-1} denotes the history of Z up to time $(t-1)$. This equation indicates that the IRFs are computed as the difference between the two conditional expectations. The first term is a forecast of Z_{t+h} at the forecast origin t conditional on a monetary policy shock μ_{MP}, while the second term is the baseline forecast (i.e. the forecast conditional on a zero monetary shock). Note that when calculating the impulse responses, I do not take into account the drift of the VAR coefficients over the impulse response horizon (during t to $t+h$).

Once the impulse responses of F_t and R_t are obtained, it is straightforward to compute the impulse responses of the observed variables X_t through the observation equation:

$$\begin{pmatrix} (IRF)^{X_1}_{t,h} \\ \vdots \\ (IRF)^{X_N}_{t,h} \end{pmatrix} = \begin{pmatrix} \Lambda^{11} & \cdots & \Lambda^{1K} & \Psi^{11} \\ \vdots & \ddots & \vdots & \vdots \\ \Lambda^{N1} & \cdots & \Lambda^{NK} & \Psi^{N1} \end{pmatrix} (IRF)^Z_{t,h}$$

Data set The data set consists of 432 monthly series, spanning the period from 1972:01 to 2007:12. The first 36 months (1972:01-1974:12) are used as a training sample to calibrate the priors. The observed panel $X_{i,t}$ consists of 182 variables, which are made up of aggregate CPI, disaggregate CPIs and other macroeconomic/financial variables. All of the variables are listed in Tables 1.2-1.4. The macroeconomic/financial variables cover a wide range of measures including real activity, employment, price changes, asset prices, exchange rates, money aggregates and interest rates. The data sources are the Ministry of Internal Affairs and Communications[15], the Ministry of Economy, Trade and Industry[16], the Bank of Japan[17], and Global Financial Data[18]. It should be also mentioned that the data set includes 91 disaggregate prices. As for these prices, the data

[15] http://www.stat.go.jp
[16] http://www.meti.go.jp
[17] https://www.boj.or.jp
[18] https://www.globalfinancialdata.com/index.html

are collected at the highest level of disaggregation[19]. The whole dataset provide comprehensive information set which the central bank analyzes in formulating monetary policies. It is, in principle, ideal to use data used by the BoJ at the time of policymaking (i.e. real-time data instead of fully-revised data). However, I assume that this issue has little impact on a precision of analysis[20].

Following common works in the literature (e.g., Stock and Watson (2005), Banbura et al. (2007)), the data series are transformed. Most of the series are transformed by using the first difference in the logs of the seasonally adjusted series by Census X13 ARIMA. However, such transformation is not applied to the variables already expressed in rates. The transformation applied to each variable is indicated in Tables 1.2-1.4. Furthermore, these tables show whether each variable is treated as a slow-moving or fast-moving variable.

As mentioned in 1.3.1, the monetary policy instrument R_t is defined by the overnight call rate for the pre-1995 period (1975-1994), and by the shadow rate for the post-1995 period (1995-2007). Note that R_t is the only one observed factor (i.e. the only observed variable contained in Z_t). These are based on the assumption that monetary policy affects a wide range of economic variables in a pervasive manner. As for the data series of the policy instrument, most of earlier empirical works apply no data-series transformation to the monetary policy instrument. However, I use the Baxter-King bandpass filter, where the lead-lag length of the filter is set equal to 12 months, and the bandpass range spans from 6 to 96 months[21].

[19] If the data in some disaggregation levels are missing during the time-span that I analyze, I switch to the next level of aggregation so then all the sub-categories of disaggregate prices are chosen without double-counting.

[20] As mentioned in Baumeister et al. (2013), this assumption is supported by Bernanke and Boivin (2003) confirming that given the latent nature of factors, the data revision has no serious impact on estimation.

[21] As revealed in appendix 1.6.4, when the raw (non-transformed) time-series of the policy instrument is used, monetary policy effects are not identified correctly. For this reason, the bandpass filter is applied to the policy instrument.

Table 1.2: List of the 182 variables used in the dataset $X_{i,t}$ of TVP/SV-FAVAR. The table also shows whether each variable is treated as a slow- or fast-moving variable, and whether log difference is applied to its time-series.

Series	Slow/Fast	Log
[Macroeconomic/financial variables]		
(1) Index of Producer's Shipments (Aggregate)	Slow	✓
(2) Index of Producer's Shipments (Construction goods)	Slow	✓
(3) Index of Producer's Shipments (Capital goods)	Slow	✓
(4) Index of Producer's Shipments (Durable consumer goods)	Slow	✓
(5) Index of Producer's Shipments (Nondurable consumer goods)	Slow	✓
(6) Index of Producer's Shipments (Consumer goods)	Slow	✓
(7) Index of Producer's Shipments (Final demand goods)	Slow	✓
(8) Index of Producer's Shipments (Investment goods)	Slow	✓
(9) Index of Producer's Shipments (Producer goods)	Slow	✓
(10) IIP (Aggregate)	Slow	✓
(11) IIP (Construction goods)	Slow	✓
(12) IIP (Capital goods)	Slow	✓
(13) IIP (Durable consumer goods)	Slow	✓
(14) IIP (Nondurable consumer goods)	Slow	✓
(15) IIP (Consumer goods)	Slow	✓
(16) IIP (Final demand goods)	Slow	✓
(17) IIP (Investment goods)	Slow	✓
(18) IIP (Producer goods)	Slow	✓
(19) Retail turnover	Slow	✓
(20) Total retail trade	Slow	✓
(21) Index of total worked hours	Slow	
(22) New job offers (1000 persons)	Slow	✓
(23) New job offers (1000 persons, part-time)	Slow	✓
(24) Regular employment index	Slow	✓
(25) Unemployment rate	Slow	
(26) Employment (Private sector)	Slow	✓
(27) Employment (All industries w/o construction)	Slow	✓
(28) Employment (Construction)	Slow	✓
(29) Employment (Agriculture)	Slow	✓
(30) Employment (Services)	Slow	✓
(31) Wage index (All industries)	Slow	✓
(32) Wage index (Manufacturing)	Slow	✓
(33) Imports of goods	Slow	✓
(34) Export of goods	Slow	✓
(35) Imports: Value goods	Slow	✓
(36) Export: Value goods	Slow	✓
(37) Total floor area of new housing (total)	Slow	✓
(38) Total floor area of new housing (owned)	Slow	✓
(39) Total floor area of new housing (rented)	Slow	✓
(40) Total floor area of new housing (for sale)	Slow	✓
(41) Total number of new housing (total)	Slow	✓
(42) Total number of new housing (owned)	Slow	✓
(43) Total number of new housing (rented)	Slow	✓
(44) Total number of new housing (for sale)	Slow	✓
(45) Composite leading indicators	Slow	✓
(46) TOPIX	Fast	✓
(47) Nikkei 225 stock average	Fast	✓
(48) Tokyo SE second section composite	Fast	✓
(49) TOPIX (Fishing and forestry)	Fast	✓
(50) TOPIX (Mining)	Fast	✓
(51) TOPIX (Construction)	Fast	✓
(52) TOPIX (Foods)	Fast	✓
(53) TOPIX (Textiles)	Fast	✓
(54) TOPIX (Machinery)	Fast	✓
(55) TOPIX (Transport equipment)	Fast	✓
(56) TOPIX (Land transportation)	Fast	✓
(57) TOPIX (Marine transportation)	Fast	✓
(58) TOPIX (Air transportation)	Fast	✓
(59) TOPIX (Communications)	Fast	✓
(60) TOPIX (Electricity and gas)	Fast	✓

Table 1.3: List of the 182 variables used in the dataset $X_{i,t}$ of TVP/SV-FAVAR (Cont'd). The table also shows whether each variable is treated as a slow- or fast-moving variable, and whether log difference is applied to its time-series.

Series	Slow/Fast	Log
(61) TOPIX (Services)	Fast	✓
(62) Exchange rate (JPY per USD)	Fast	✓
(63) Exchange rate (JPY per GBP)	Fast	✓
(64) JPY effective exchange rate (nominal)	Fast	✓
(65) JPY effective exchange rate (real)	Fast	✓
(66) Money stock (M1)	Fast	✓
(67) Money stock (M2+CD)	Fast	✓
(68) Monetary base	Fast	✓
(69) Index of producer's inventory (Aggregate)	Fast	✓
(70) Index of producer's inventory (Construction goods)	Fast	✓
(71) Index of producer's inventory (Capital goods)	Fast	✓
(72) Index of producer's inventory (Durable consumer goods)	Fast	✓
(73) Index of producer's inventory (Nondurable consumer goods)	Fast	✓
(74) Index of producer's inventory (Consumer goods)	Fast	✓
(75) Index of producer's inventory (Final demand goods)	Fast	✓
(76) Index of producer's inventory (Investment goods)	Fast	✓
(77) Index of producer's inventory (Producer goods)	Fast	✓
(78) Discount rate	Fast	
(79) Long prime rate	Fast	
(80) Average lending rate	Fast	
(81) Nikko convertible bond price index	Fast	✓
(82) Corporate bond yield	Fast	
(83) Government bond yield (10yr)	Fast	
(84) Government bond yield (7yr)	Fast	
(85) Private bills discount rate (3months)	Fast	
(86) Tbill yield (3months)	Fast	
[CPIs]		
(87) CPI (Aggregate)	Slow	✓
(88) CPI (Aggregate, less fresh food)	Slow	✓
(89) CPI (Aggregate, less imputed rent)	Slow	✓
(90) CPI (Aggregate, less imputed rent and fresh food)	Slow	✓
(91) CPI (Aggregate, less food and energy)	Slow	✓
(92) CPI (Cereals)	Slow	✓
(93) CPI (Fish and seafood)	Slow	✓
(94) CPI (Meats)	Slow	✓
(95) CPI (Dairy products and eggs)	Slow	✓
(96) CPI (Vegetables and seaweeds)	Slow	✓
(97) CPI (Fruits)	Slow	✓
(98) CPI (Oils, fats and seasonings)	Slow	✓
(99) CPI (Cakes and candies)	Slow	✓
(100) CPI (Cooked food)	Slow	✓
(101) CPI (Beverages)	Slow	✓
(102) CPI (Alcoholic beverages)	Slow	✓
(103) CPI (Meals outside the home)	Slow	✓
(104) CPI (House rent, private)	Slow	✓
(105) CPI (House rent, public)	Slow	✓
(106) CPI (Imputed rent)	Slow	✓
(107) CPI (Materials for repairs and maintenance)	Slow	✓
(108) CPI (Service charges for repairs and maintenance)	Slow	✓
(109) CPI (Electricity)	Slow	✓
(110) CPI (Gas)	Slow	✓
(111) CPI (Kerosene)	Slow	✓
(112) CPI (Water and sewerage charges)	Slow	✓
(113) CPI (Durable goods assisting housework)	Slow	✓
(114) CPI (Heating and cooling appliances)	Slow	✓
(115) CPI (General furniture)	Slow	✓
(116) CPI (Interior furnishings)	Slow	✓
(117) CPI (Bedding)	Slow	✓
(118) CPI (Tableware)	Slow	✓
(119) CPI (Kitchen utensils)	Slow	✓
(120) CPI (Other domestic utensils)	Slow	✓

Table 1.4: List of the 182 variables used in the dataset $X_{i,t}$ of TVP/SV-FAVAR (Cont'd). The table also shows whether each variable is treated as a slow- or fast-moving variable, and whether log difference is applied to its time-series.

Series	Slow/Fast	Log
(121) CPI (Facial tissue and rolled toilet paper)	Slow	✓
(122) CPI (Detergent)	Slow	✓
(123) CPI (Other nondurable goods)	Slow	✓
(124) CPI (Japanese clothing)	Slow	✓
(125) CPI (Men's clothing)	Slow	✓
(126) CPI (Women's clothing)	Slow	✓
(127) CPI (Children's clothing)	Slow	✓
(128) CPI (Men's shirts and sweaters)	Slow	✓
(129) CPI (Women's shirts and sweaters)	Slow	✓
(130) CPI (Children's shirts and sweaters)	Slow	✓
(131) CPI (Men's underwear)	Slow	✓
(132) CPI (Women's underwear)	Slow	✓
(133) CPI (Children's underwear)	Slow	✓
(134) CPI (Footwear)	Slow	✓
(135) CPI (Other clothing)	Slow	✓
(136) CPI (Laundry charges)	Slow	✓
(137) CPI (Dry cleaning charges)	Slow	✓
(138) CPI (Footwear repair charges)	Slow	✓
(139) CPI (Medicines and health fortification)	Slow	✓
(140) CPI (Medical supplies and appliances)	Slow	✓
(141) CPI (Medical services)	Slow	✓
(142) CPI (Public transportation)	Slow	✓
(143) CPI (Automobiles)	Slow	✓
(144) CPI (Bicycles)	Slow	✓
(145) CPI (Automotive maintenance)	Slow	✓
(146) CPI (Communication)	Slow	✓
(147) CPI (PTA membership fees at elementary school)	Slow	✓
(148) CPI (PTA membership fees at junior high school)	Slow	✓
(149) CPI (Junior high school fees, private)	Slow	✓
(150) CPI (High school fees, public)	Slow	✓
(151) CPI (High school fees, private)	Slow	✓
(152) CPI (College and university fees, national)	Slow	✓
(153) CPI (College and university fees, private)	Slow	✓
(154) CPI (Vocational school fees)	Slow	✓
(155) CPI (Kindergarten fees, public)	Slow	✓
(156) CPI (Kindergarten fees, private)	Slow	✓
(157) CPI (School textbooks and ref books for study)	Slow	✓
(158) CPI (Tutorial fees)	Slow	✓
(159) CPI (Recreational durable goods)	Slow	✓
(160) CPI (Stationery)	Slow	✓
(161) CPI (Sporting goods)	Slow	✓
(162) CPI (Toys)	Slow	✓
(163) CPI (Cut flowers)	Slow	✓
(164) CPI (Other recreational goods)	Slow	✓
(165) CPI (Newspapers)	Slow	✓
(166) CPI (Magazines)	Slow	✓
(167) CPI (Books)	Slow	✓
(168) CPI (Hotel charges)	Slow	✓
(169) CPI (Package tours)	Slow	✓
(170) CPI (Lesson fees)	Slow	✓
(171) CPI (Charges for TV license)	Slow	✓
(172) CPI (Admission and game charges)	Slow	✓
(173) CPI (Other recreational services)	Slow	✓
(174) CPI (Personal care services)	Slow	✓
(175) CPI (Toilet utensils)	Slow	✓
(176) CPI (Soap and others)	Slow	✓
(177) CPI (Cosmetics)	Slow	✓
(178) CPI (Bags)	Slow	✓
(179) CPI (Watches and rings)	Slow	✓
(180) CPI (Other personal effects)	Slow	✓
(181) CPI (Tobacco)	Slow	✓
(182) CPI (Other miscellaneous)	Slow	✓

1.3.4 Model selection

The number of factors is optimized by using information criteria developed by Bai and Ng (2002):

$$\begin{cases} PC_{p1}(K) &= V(K,\hat{F}^K) + K\hat{\sigma}^2(\frac{N+T}{NT})\ln(\frac{NT}{N+T}) \\ PC_{p2}(K) &= V(K,\hat{F}^K) + K\hat{\sigma}^2(\frac{N+T}{NT})\ln C_{NT}^2 \\ PC_{p3}(K) &= V(K,\hat{F}^K) + K\hat{\sigma}^2(\frac{\ln C_{NT}^2}{C_{NT}^2}) \\ IC_{p1}(K) &= \ln(V(K,\hat{F}^K)) + K(\frac{N+T}{NT})\ln(\frac{NT}{N+T}) \\ IC_{p2}(K) &= \ln(V(K,\hat{F}^K)) + K(\frac{N+T}{NT})\ln C_{NT}^2 \\ IC_{p3}(K) &= \ln(V(K,\hat{F}^K)) + K(\frac{\ln C_{NT}^2}{C_{NT}^2}) \end{cases}$$

where $V(K,\hat{F}^K) = N^{-1}\Sigma_{i=1}^N \hat{\sigma}_i^2$ with $\hat{\sigma}_i^2 = \hat{e}_i'\hat{e}_i/T$, $\hat{\sigma}^2 = \frac{1}{NT}\Sigma_{i=1}^N\Sigma_{t=1}^T\hat{E}(e_{it})^2$, and $C_{NT} = \min\{\sqrt{N},\sqrt{T}\}$. Note also that K represents the number of factors. Table 1.5 reveals that all the criteria except for IC_{p3} are minimized in the case of $K = 3$[22]. This clearly indicates that $K = 3$ is the optimized result.

Table 1.5: Result for the information criteria defined by Bai and Ng (2002): K is the number of factors.

K	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
2	6.11e-01	6.16e-01	5.94e-01	-4.87e-01	-4.77e-01	-5.55e-01
3	6.03e-01	6.07e-01	5.92e-01	-5.02e-01	-4.95e-01	-5.46e-01

1.4 Analysis results

This section reports the analysis results obtained from the estimated FAVAR.

1.4.1 Factors and volatility

Fig. 1.4 displays the estimated factors, monetary policy instrument, and stochastic volatility of shocks to each transition equation. Factor 1 is

[22] Throughout this study, I do not consider the models with $K \geq 4$. This is because when $K \geq 4$, the computation time explodes and the model's estimation performance becomes unstable.

strongly correlated with financial variables[23]. Volatility $h_{1,t}$ temporarily increases around 1990, and it is clearly associated with the periods of bubble generation and its collapse. Factor 2 is correlated with IIP (real output) with correlation coefficient of 0.84. Volatility $h_{2,t}$ is larger during the 1990s-2000s than the one during the 1970s-1980s, and this is consistent with the feature of IIP's volatility[24]. Factor 3 tends to be correlated with the aggregate CPI, where the correlation coefficient is 0.66. Volatility $h_{3,t}$ significantly decreases during the 1970s to early 1980s, which is in line with the feature of CPI's volatility[25]. As for volatility $h_{4,t}$, it rises around 1980 and 1985, and this seems to reflect the temporary increase of the call rate around those periods. During the periods of the long-term recession and (near-)zero interest rate regime (1990s-2000s), $h_{4,t}$ is much smaller than during the earlier periods (see also Fig. 1.5).

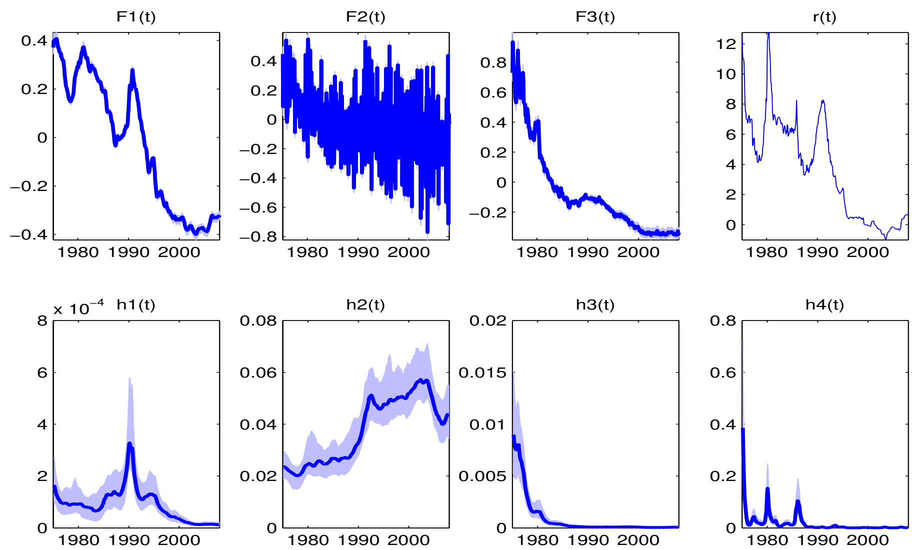
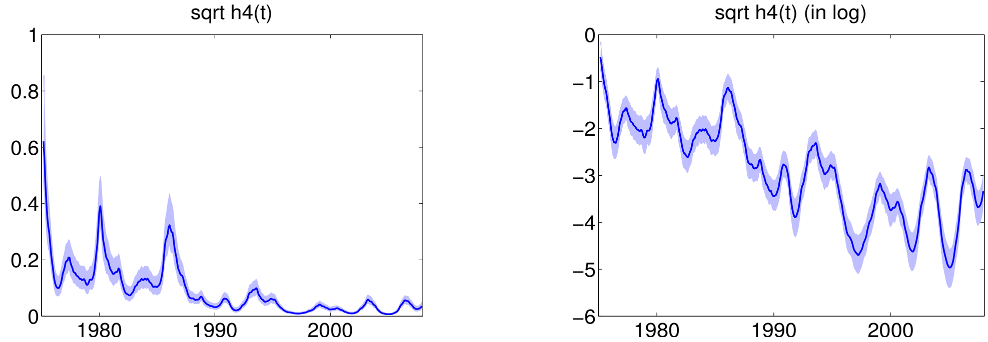


Figure 1.4: Estimated factors ($F_{1,t}$, $F_{2,t}$, $F_{3,t}$), monetary policy instrument (R_t) and stochastic volatility ($h_1(t)$, $h_2(t)$, $h_3(t)$, $h_4(t)$): The point estimate (blue solid line) is shown with the 16-th and 84-th percentiles (shaded area).

[23]It is correlated with interest rates such as the government bond yield, with correlation coefficient of larger than 95%.

[24]As for the IIP's first log difference, the standard deviation is 0.96 (1970s), 0.77 (1980s), 1.06 (1990s), and 1.14 (2000s).

[25]As for the CPI's first log difference, the standard deviation is 1.03 (1970s), 0.75 (1980s), 0.56 (1990s), and 0.36 (2000s).



(a) $\sqrt{h_4(t)}$ (b) $log\sqrt{h_4(t)}$

Figure 1.5: $\sqrt{h_4(t)}$ and $log\sqrt{h_4(t)}$: The point estimate (blue solid line) is shown with the 16-th and 84-th percentiles (shaded area).

1.4.2 Time variations in monetary policy transmission

Impulse responses to monetary policy shock Fig. 1.6 displays the impulse response of the aggregate CPI to a tightening monetary policy shock. Throughout this study, the monetary policy shock is defined as an increase of 100 basis points of the monetary policy instrument R_t. During the late 1990s and 2000s, the policy shock brings a decrease of 0.7-1.0 % after 48 months (h=48), which is a larger impact as compared with the earlier periods. Time variation in the response is exhibited in Fig. 1.6(c). As a variation pattern for the (absolute) response strength, it slightly decreases during the late 1970s, becomes relatively flat in the 1980s, rapidly increases during the 1990s, and subsequently becomes flat again in the 2000s.

Following the approach by Cogley et al. (2010), Fig. 1.6(d) examines the relative importance of time variation in the impulse response. It plots the joint posterior distribution of the response at two points in time (t_1, t_2) (=(1975,2005) or (1985,2005)), and shifts of the distribution away from the 45° line represent a change in the response strength across time. Fraction of the joint distribution above the 45° line is found as:

$$\text{Prob}(\Delta irf > 0) = \begin{cases} 0.22 & (\text{for } (t_1, t_2) = (1975{,}2005)) \\ 0.07 & (\text{for } (t_1, t_2) = (1985{,}2005)) \end{cases}$$

This parameter corresponds to the p-value for rejecting a hypothesis that there is no time variation in the price responses. The result p = 0.07

corresponds to a relatively significant evidence of time variation.

As mentioned in Ellis et al. (2014), Fig. 1.6 can not necessarily illustrate a time variation in the transmission of monetary policy shock in an adequate manner. This is because even if I use the policy shock with a fixed size (= 100 basis points) over the whole sample period, persistence of the monetary policy instrument R_t may change over time significantly. However, Fig. 1.7 suggests that such a concern is useless. This figure displays the impulse response of the policy instrument to the tightening policy shock, indicating that the response behavior hardly changes over time.

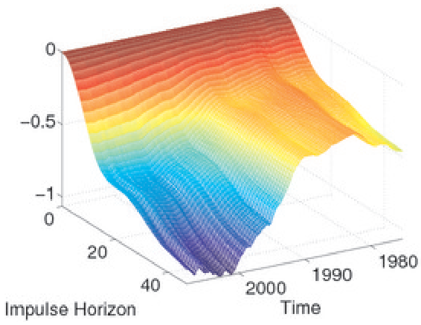
Fig. 1.8 shows the impulse response of IIP. Figs. 1.8(a) and (c) suggest that the (absolute) strength of the long-run response slightly increases over time. However, fraction of the joint distribution above the 45° line (see Fig. 1.8(d)) is 34 % in both cases of (t_1,t_2)=(1975,2005) and (1985,2005), which indicates no significant evidence of time variation.

Explaining the results In the previous paragraph, the impulse response of aggregate price becomes stronger during the early 1990s, whereas that of real output does not exhibit a significant time variation. In what follows, I examine the theoretical plausibility of this finding, by conducting a simple simulation exercise.

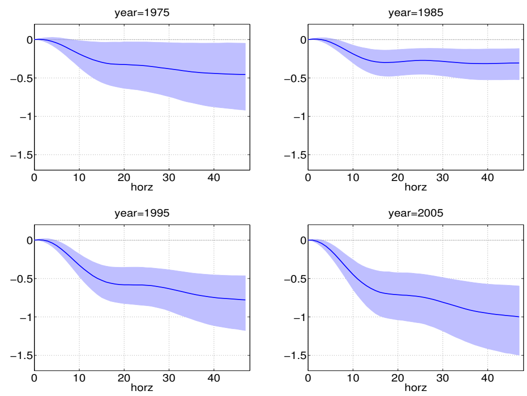
The exercise is performed by using the New Keynesian model with hybrid Phillips curve. The equation system of the model is given by:

$$
\begin{aligned}
y_t &= \gamma y_{t+1|t} + (1-\gamma) y_{t-1} - \sigma^{-1}(R_t - \pi_{t+1|t}) \\
\pi_t &= \frac{\beta}{1+\alpha\beta}\pi_{t+1|t} + \frac{\alpha}{1+\alpha\beta}\pi_{t-1} + \kappa y_t \\
R_t &= \rho R_{t-1} + (1-\rho)(\phi_\pi \pi_t + \phi_y y_t + \varepsilon_{R,t}), \quad \varepsilon_{R,t} = \rho_R \varepsilon_{R,t-1} + \tilde{\varepsilon}_{R,t}
\end{aligned}
$$

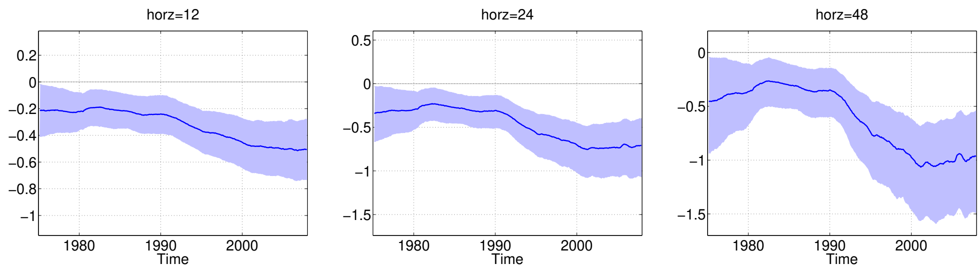
where π_t, y_t and R_t are inflation, output gap and nominal interest rate, respectively. The first equation is the intertemporal IS curve, where γ characterizes the forward-looking component and σ represents the sensitivity of output to monetary policy. The second equation is the hybrid Phillips curve, where α and κ represent inflation's persistence and sensitivity to output, respectively. All the variables in the IS and Phillips curves are expressed as log-deviations from a steady-state. The third line represents an interest rate rule, where $\tilde{\varepsilon}_{R,t}$ is a disturbance to the monetary policy rule, and its persistence is characterized by ρ_R.



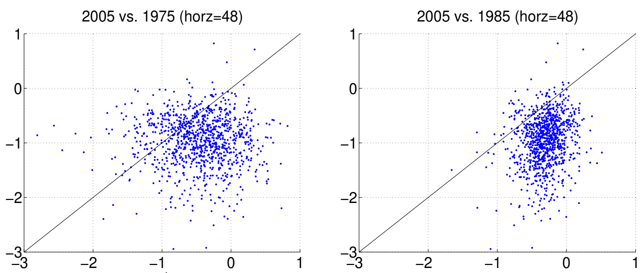
(a) Time-varying median response



(b) Response at a fixed point of time (1975/85/95/2005)

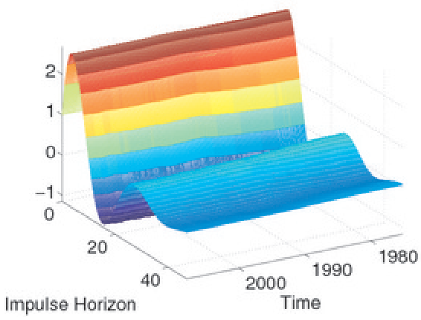


(c) Response vs. time (left : h=12, middle : h=24, right : h=48)

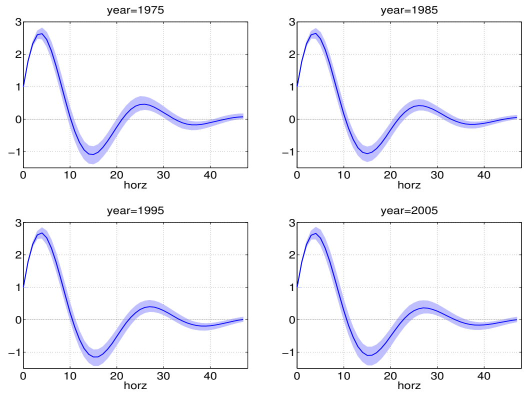


(d) Joint distribution (left : 2005 vs. 1975, right : 2005 vs. 1985)

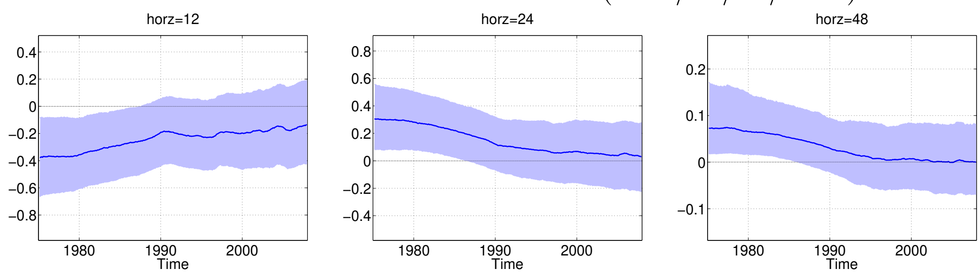
Figure 1.6: Impulse response of the aggregate CPI to a 1% increase in the monetary policy instrument: (a) time-varying median response; (b) median response (thick line) with the 16-th and 84-th percentiles (shaded area) in 1975:06, 1985:06, 1995:06 and 2005:06; (c) median response (thick line) with the 16-th and 84-th percentiles (shaded area) at an impulse horizon of 12,24 and 48 months ($h = 12, 24, 48$); (d) joint distribution of the responses at (t_1, t_2)=(1975,2005) and (1985,2005) at an impulse horizon of 48 months (h=48) (x-axis: response at time t_1, y-axis: response at time t_2).



(a) Time-varying median response

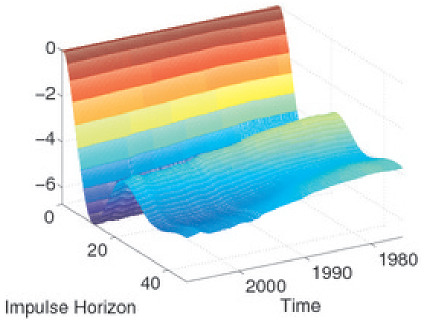


(b) Response at a fixed point of time (1975/85/95/2005)

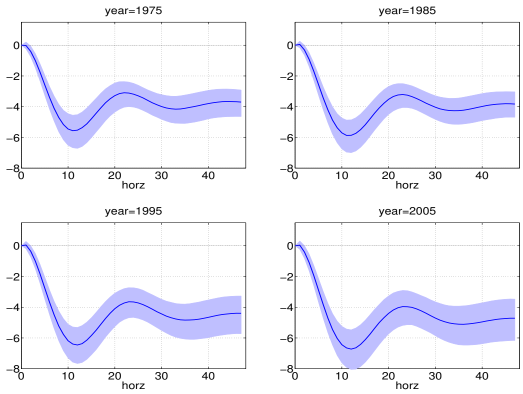


(c) Response vs. time (left : h=12, middle : h=24, right : h=48)

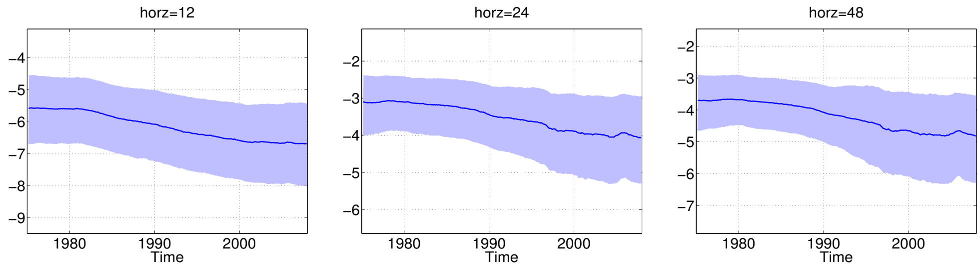
Figure 1.7: Impulse response of the monetary policy instrument to a 1% increase in the monetary policy instrument: (a) time-varying median response; (b) median response (thick line) with the 16-th and 84-th percentiles (shaded area) in 1975:06, 1985:06, 1995:06 and 2005:06; (c) median response (thick line) with the 16-th and 84-th percentiles (shaded area) at an impulse horizon of 12,24 and 48 months ($h = 12, 24, 48$).



(a) Time-varying median response



(b) Response at a fixed point of time (1975/85/95/2005)



(c) Response vs. time (left : h=12, middle : h=24, right : h=48)

(d) Joint distribution (left : 2005 vs. 1975, right : 2005 vs. 1985)

Figure 1.8: Impulse response of the IIP to a 1% increase in the monetary policy instrument: (a) time-varying median response; (b) median response (thick line) with the 16-th and 84-th percentiles (shaded area) in 1975:06, 1985:06, 1995:06 and 2005:06; (c) median response (thick line) with the 16-th and 84-th percentiles (shaded area) at an impulse horizon of 12,24 and 48 months ($h = 12, 24, 48$); (d) joint distribution of the responses at (t_1, t_2)=(1975,2005) and (1985,2005) at an impulse horizon of 48 months (h=48) (x-axis: response at time t_1, y-axis: response at time t_2).

Table 1.6: Estimated result by Benati (2008)

Parameters	1975-2006 (Full period)	1983-2006 (Post-Great-Inflation)
γ	1.0	1.0
σ	25.3	9.9
α	0.67	0.46
κ	0.037	0.052
ρ	0.94	0.87
ϕ_π	0.99	1.24
ϕ_y	1.42	0.79
ρ_R	0.22	0.09

As for calibration, the estimates of Benati (2008) are used. With Japan's data, he estimates the above model for two sample periods: *the full sample period* (1957-2006) and *the post-Great-Inflation period* (1983-2006). Table 1.6 displays his estimates. Throughout the exercise, his estimates for the full period are used as a benchmark setup for the calibration. In addition, I also use other eight alternative setups, where in each setup, one of the eight parameters is calibrated by the estimate for the post-Great-Inflation period.

The solution form of the model is obtained by converting the equation system of the data-generating process to VAR form[26]. Then, by simulating the solution form, I calculate the impulse responses of the economic variables to monetary policy shock at the theoretical level. Figs. 1.9 and 1.10 display the obtained results. In each figure, black (solid) line and red (dashed) line represent the responses with the baseline and alternative parameter setups, respectively. The figures show that the change in the parameter (κ or ϕ_y) makes the response of inflation stronger, whereas those of output and interest rate exhibit little variation. The feature of these changes is consistent with the observation in Figs. 1.6-1.8. In other words, this implies that the observation in Figs. 1.6-1.8 can be possibly explained by a transition in the dynamics of the underlying economy.

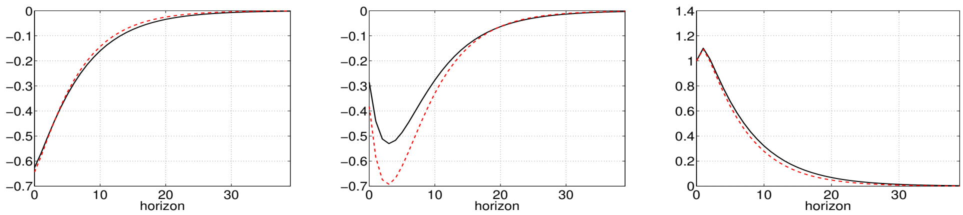
As for the responses with the other alternative setups, see appendix 1.6.5.

[26]The form is written by

$$\widetilde{X}_t = A\widetilde{X}_{t-1} + Bu_t$$

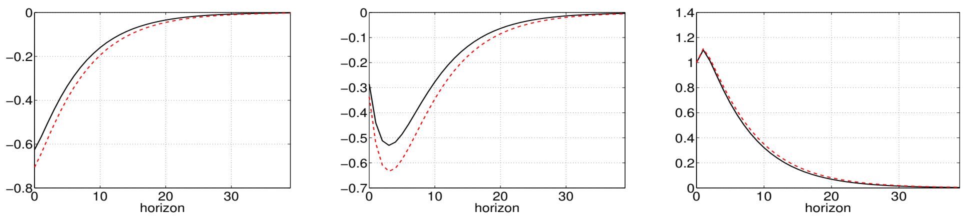
where $\widetilde{X}_t$ and u_t are a vector of endogenous variables and structural shocks, respectively. This solution form is often called the *Sims' canonical form.* As for how to obtain the matrices A and B, see Sims (2002).

In each case, the impulse responses display some changes in comparison to those with the baseline setup. However, the feature of the change is not necessarily in agreement with the one observed in Figs. 1.6-1.8.



(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.9: Impulse responses to a tightening monetary policy shock of 100 basis points when κ is changed (κ: 0.037$\rightarrow$ 0.052): Black solid line and red dashed line represent the responses before and after changing κ, respectively.



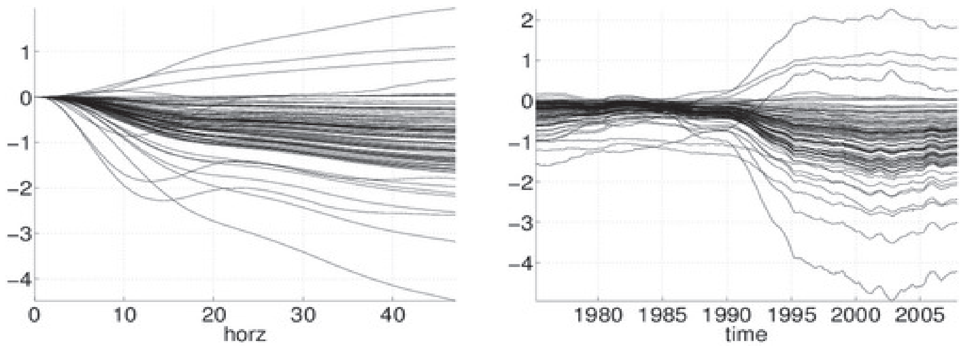
(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.10: Impulse responses to a tightening monetary policy shock of 100 basis points when ϕ_y is changed (ϕ_y: 1.42$\rightarrow$ 0.79): Black solid line and red dashed line represent the responses before and after changing ϕ_y, respectively.

1.4.3 Monetary policy transmission at disaggregate level

As described in 1.3.3, the dataset used in the analysis includes 91 disaggregate prices. In what follows, I examine monetary policy transmission to the disaggregate prices, to see whether the monetary policy effects change at not only aggregate but also disaggregate level.

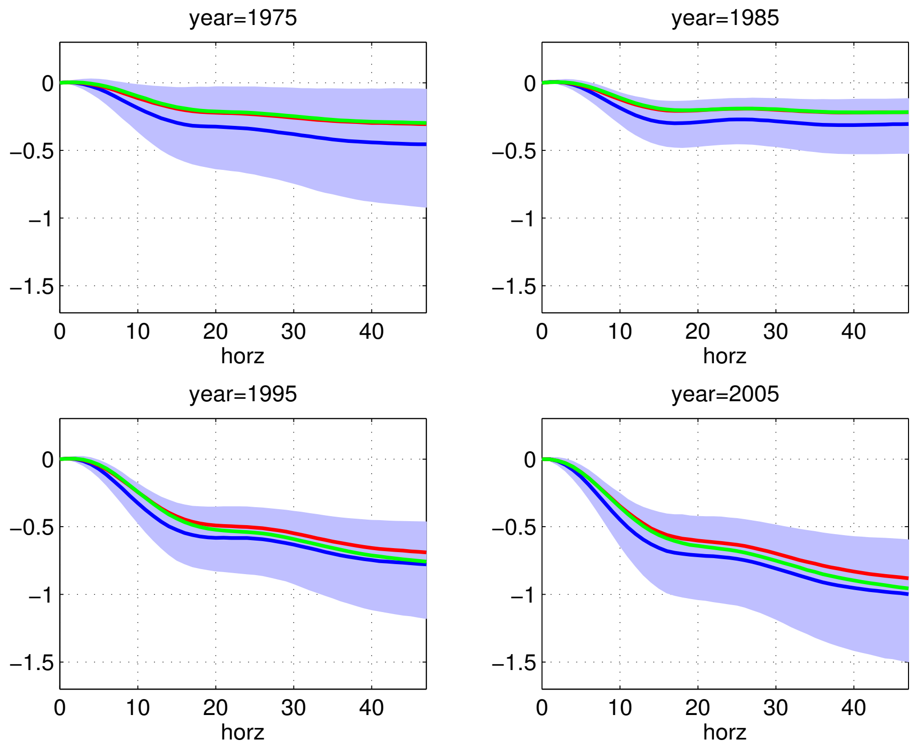
Typical responses of disaggregate prices Fig. 1.11 displays the impulse response of the 91 disaggregate prices to a tightening monetary policy shock of 100 basis points. To understand the typical response of the disaggregate prices, Fig. 1.12 shows the weighted mean (red line) and unweighted mean (green line) responses of the 91 prices. The aggregate price response is also plotted as a reference (blue line), along with the 16-th and 84-th percentiles of the posterior distribution. The response pattern of the green and red lines is very similar to that of the aggregate price. As for the response magnitude, those two responses do not significantly deviate from the estimate of aggregate CPI. This suggests that the aggregate price well reflects the typical behavior of price response at disaggregate level.



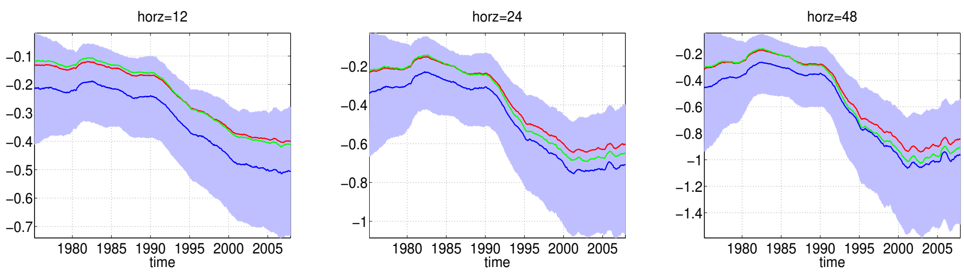
(a) Response in 2005:06 (b) Response vs. time (h=48)

Figure 1.11: Impulse response of the 91 disaggregate prices to a 1% increase in the monetary policy instrument

Heterogeneity of price responses Fig. 1.11 indicates an existence of heterogeneity in the responses of 91 disaggregate prices. To discuss a statistical significance of this heterogeneity, Fig. 1.13 examines the weighted cross sectional distribution of the impulse response of the 91 disaggregate prices (at h=48). Note that the distributions are obtained from the last $(M - M_0)$ iterations of the Bayesian estimation. In the figure, green lines represent the mean (solid) and one standard deviation (dashed). Red lines show the mean and one standard deviation obtained from the posterior distribution of the weighted mean response of the 91 prices (the point estimate and the estimation uncertainty of the weighted mean response). Bandwidth of the green dashed lines represents a degree of heterogeneity in the disaggregate price responses, while that of the red dashed lines indicates an estimation uncertainty of the typical response of the disaggregate prices. Fig. 1.13 shows that the former is about twice as wide as the latter in 1975 and 1985, and more than three times wider



(a) Response at a fixed point of time (1975/85/95/2005)



(b) Response vs. time (h=12,24,48)

Figure 1.12: Weighted-mean (red line) and unweighted-mean (green line) responses of the 91 disaggregate prices, and the median response (blue line) of the aggregate CPI with the 16-th and 84-th percentiles (shaded area): (a) response at a fixed point of time (1975:06, 1985:06, 1995:06 and 2005:06); (b) response at an impulse horizon of 12, 24 and 48 months (h=12,24 and 48).

in 1995 and 2005. This implies that the spread of the cross sectional distribution cannot explained by the estimation uncertainty, and in this sense, the observed heterogeneity in the disaggregate price responses can be regarded as statistically significant.

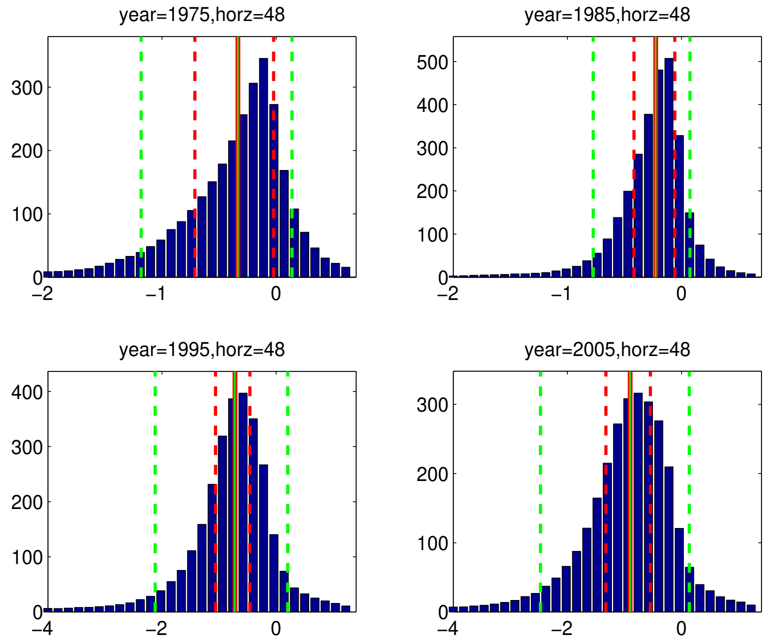
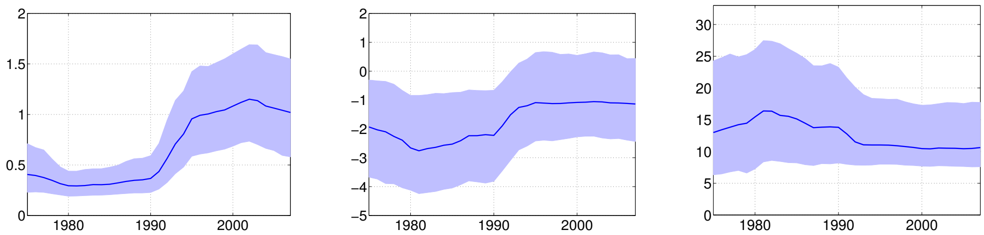


Figure 1.13: Weighted cross-sectional distribution of the impulse response of the 91 disaggregate prices in 1975:06, 1985:06, 1995:06 and 2005:06 at an impulse horizon of 48 months (h=48): Green lines represent the mean (solid) and one standard deviation (dashed) of the cross-sectional distribution. Red lines are the mean (solid) and one standard deviation (dashed) of the posterior distribution of the weighted-mean response of the 91 disaggregate prices.

Fig. 1.14 shows the time variation in the standard deviation of the weighted cross-sectional distribution of the 91 price responses. It rapidly increases during the early 1990s, indicating a significant change in dispersion of the price responses. As a further investigation, the figure also examines skewness and kurtosis. However, a remarkable variation is not necessarily observed.

1.5 Conclusion

In this chapter, I investigate time variations in the monetary policy effects from the 1970s to the 2000s in Japan, where the main focus is whether



(a) Std (b) Skewness (c) Kurtosis

Figure 1.14: Time variation in (a) standard deviation (std), (b) skewness and (c) kurtosis of the weighted cross-sectional distribution of the 91 disaggregate prices at an impulse horizon of 48 months (h=48): The blue solid line indicates the median estimate and the shaded area represents the 16-th and 84-th percentiles.

and how the policy effects change due to the bubble burst (around 1990-91) and through the (near-)zero interest rate period (since 1995). The analysis is made up of two steps. First, the shadow rate is estimated with the non-linear term structure model SRTSM, where the shadow rate represents policy stance of the monetary authorities during the (near-)zero interest rate period. Then, using the obtained shadow rate as monetary policy instrument, the second step is devoted to estimating the time-varying FAVAR. The estimated impulse response of the real output does not exhibit a clear variation across the whole sample period. However, the response of the aggregate price becomes stronger during the early 1990s, and the timing of this change is clearly associated with the bubble burst. Regarding the policy effects during the (near-)zero interest rate regime, no significant evidence of their time variation is observed. A series of simulation works with the New Keynesian model suggests that the observed variation in monetary policy transmission can be possibly explained by shifts in the underlying dynamics in the economy, such as a transition in the monetary policy rule. As for the price response to monetary policy shock, I also examine disaggregate level. The estimated responses of the 91 disaggregate prices indicate the significant time variations in terms of not only response strength but also cross-sectional dispersion.

1.6 Appendices

1.6.1 Approximation to ρ_1^Q

As explained in 1.2.2, the forward rate $f_{m,m+1,t}$ can be written as:

$$f_{m,m+1,t} = r_{min} + \sigma_m(\tilde{\theta}) \cdot g\left(\frac{a_m(\tilde{\theta}) + b_m(\tilde{\theta})X_t - r_{min}}{\sigma_m(\tilde{\theta})}\right) \tag{1.13}$$

where $\theta = \{\delta_0, \rho^Q, Q\}$. As can be understood from Wu and Xia (2016), when m is sufficiently large (for example, $m \geq 60$ etc.), the following holds:

$$g\left(\frac{a_m(\tilde{\theta}) + b_m(\tilde{\theta})X_t - r_{min}}{\sigma_m(\tilde{\theta})}\right) \simeq \frac{a_m(\tilde{\theta}) + b_m(\tilde{\theta})X_t - r_{min}}{\sigma_m(\tilde{\theta})}$$

Under this situation, Eq. (1.13) becomes

$$\begin{aligned}
f_{m,m+1,t} &\simeq r_{min} + a_m(\tilde{\theta}) + b_m(\tilde{\theta})X_t - r_{min} \\
&= a_m(\tilde{\theta}) + b_m(\tilde{\theta})X_t \\
&= \delta_0 + \delta_1'(\sum_{j=0}^{m-1}(\rho^Q)^j)\mu^Q + \delta_1'(\rho^Q)^m X_t \\
&= \delta_0 + \delta_1'(\rho^Q)^m X_t \qquad \text{(from} \quad \mu^Q = 0) \\
&= \delta_0 + [1 \;\; 1 \;\; 0] \begin{pmatrix} \rho_1^Q & 0 & 0 \\ 0 & \rho_2^Q & 1 \\ 0 & 0 & \rho_2^Q \end{pmatrix}^m \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{pmatrix} \\
&= \delta_0 + (\rho_1^Q)^m X_{1,t} + (\rho_2^Q)^m X_{2,t} + m \cdot (\rho_2^Q)^{m-1} X_{3,t}
\end{aligned} \tag{1.14}$$

Then, as shown in Fig. 1.15, the forward rate curve exhibits the following features:

- In large-m region, $f_{m,m+1,t}$ becomes almost constant. ($f_{m,m+1,t}|_{m:\text{large}}$ hardly depends on m.)
- The level of $f_{m,m+1,t}|_{m:\text{large}}$ varies over time t.

If we assume that both of $|\rho_1^Q|$ and ρ_2^Q are less than 1, Eq. (1.14) cannot satisfy these features. Therefore, either of $|\rho_1^Q|$ or $|\rho_2^Q|$ needs to be one. Thus, from $(1 \geq)|\rho_1^Q| > |\rho_2^Q|$[27], $|\rho_1^Q|$ needs to be one.

[27] See (iv) in 1.2.3.

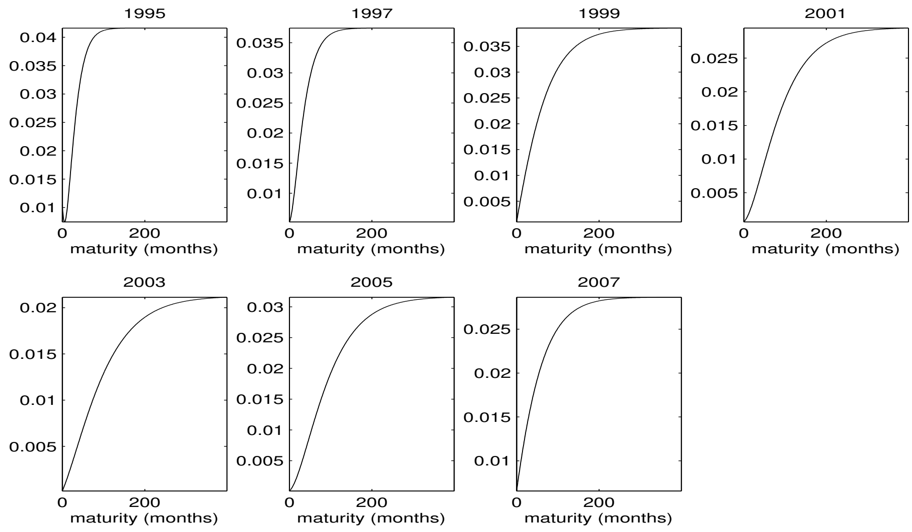


Figure 1.15: Forward rate curves in 1995:06, 1997:06, 1999:06, 2001:06, 2003:06, 2005:06 and 2007:06: The curves are obtained by extrapolating the data with the Svensson model (see Svensson (1995)).

1.6.2 Estimation procedure of time-varying FAVAR

In the time-varying FAVAR model with stochastic volatility, the parameters to be estimated are: Γ (Factor loadings); R (Covariance matrix for X_t); $\{\beta_t\}_{t=1}^T$ (VAR coefficients in the transition equation); Q (Covariance matrix for β_t); $\{a_{ij,t}\}_{t=1}^T$ (Off-diagonal elements of A_t); D (Covariance matrix for $a_{ij,t}$); $\{h_{i,t}\}_{t=1}^T$ (Diagonal elements of H_t); g_i (Variance of $\ln(h_{i,t})$); and $\{F_t^j\}_{t=1}^T$ (Factors). The estimation procedure is summarized below.

Step1: Set the priors and starting values for the model parameters. (For detail, see 1.6.3.)

(Step1a) Parameters in the transition equation

- Starting value of $\{F_t\}_{t=1}^T$
- Prior for F_0 (= initial value of F_t)
- Prior and starting value for Q

(Step1b) Parameters in the RW process for $a_{ij,t}$

- Starting value of $\{a_{ij,t}\}_{t=1}^T$
- Prior for $a_{ij,0}$ (= initial value of $a_{ij,t}$)
- Prior and starting value for D

(Step1c) Parameters in the observation equation

- Prior for Γ
- Prior and starting value for R

(Step1d) Parameters in the RW process for $\ln(h_{i,t})$

- Starting value for $\{\ln(h_{i,t})\}_{t=1}^{T}$
- Prior for $\ln(h_{i,0})$ (= initial value of $\ln(h_{i,t})$)
- Prior and starting value for g_i

Step2: Given R and Z_t, draw Γ.

Step3: Given Γ and Z_t, draw R.

Step4: Given Z_t, Q, $a_{ij,t}$ and $h_{i,t}$, draw β_t.

Step5: Given β_t, draw Q.

Step6: Given Z_t, β_t, $h_{i,t}$ and D, draw $a_{ij,t}$.

Step7: Given $a_{ij,t}$, draw D.

Step8: Given Z_t, β_t and g_i, draw $h_{i,t}$.

Step9: Given $h_{i,t}$, draw g_i.

Step10: Given Γ, R, β_t, $a_{ij,t}$ and $h_{i,t}$, draw F_t.

Step11: Iterate steps 2 to 10 by M times. When M and M_0 are sufficiently large but $M > M_0$, the marginal posterior distribution of each parameter can be approximately obtained from the last $(M - M_0)$ iterations.

1.6.3 Priors and starting values

As for the priors and starting values of the model parameters, I adopt those used in Baumeister et al. (2013) and Ellis et al. (2014).

Parameters in the transition equation (Step 1a)

- The starting values for $\{F_t\}_{t=1}^{T}$ are determined by the principal components. The principal components are also used as the central values for the prior of the initial values ($\equiv \beta_{0|0}$). The prior covariance $p_{0|0}$ (i.e. uncertainty of the initial values) is set to be the identity matrix.

- Prior and starting value for Q: The prior for Q is specified by the inverse Wishart distribution:

$$Q \sim iW(Q_0, T_0)$$

where $Q_0 = Var[\widehat{\beta}^{OLS}] \times \tau \times T_0$. $\widehat{\beta}^{OLS}$ represents the OLS estimate, which is obtained by OLS estimation over the training sample via a fixed-coefficient VAR made up of the principal components $(PC)_t$ and the monetary policy instrument R_t. T_0 is a length of the training sample. τ is set equal to 3.5×10^{-4}. Furthermore, Q_0 is used as the starting value for Q.
- VAR coefficients $\{\beta_t\}_{t=1}^{T}$: The central value and covariance of the prior for the initial value ($\beta_{0|0}$ and $p_{0|0}$) are determined via the OLS estimate over the training sample. $\beta_{0|0}$ is also used as the starting value for $\{\beta_t\}_{t=1}^{T}$.

Parameters in the RW process of $a_{ij,t}$ (Step 1b)

- Prior and starting value for $\{a_{ij,t}\}_{t=1}^{T}$: The prior for the initial value is

$$a_{ij,0} \sim N(\widehat{a}_{ij}^{OLS}, \widehat{V}_{ij}^{a})$$

where $\widehat{a}_{ij}^{OLS}$ represents the off-diagonal components of the Cholesky decomposition of $\widehat{v}^{OLS}$ with each row scaled by the corresponding element on the diagonal. $\widehat{V}_{ij}^{a}$ is set equal to $10 \times |\widehat{a}_{ij}^{OLS}|$. $\widehat{a}_{ij}^{OLS}$ is also used as the starting value for $\{a_{ij,t}\}_{t=1}^{T}$.
- Prior and starting value for D: Let D_i denote the covariance matrix of the i-th row of A. The prior of D_i is assumed to be the inverse Wishart distribution:

$$D_i \sim iW(\overline{D}_i, K_i)$$

where $\overline{D}_i$ is a corresponding element of $\widehat{a}^{OLS}$ multiplied by 10^{-3}. Moreover, $\overline{D}_i$ is used as a starting value for D_i.

Parameters in the observation equation (Step 1c)

- Prior for Γ: The prior for Γ_i (i-th row of Γ) is assumed to be normal:

$$\Gamma_i \sim N(f_\Gamma, V_\Gamma)$$

where f_Γ is $1 \times (K+1)$ vector and V_Γ is $(K+1) \times (K+1)$ matrix. They are set to be a zero vector and identity matrix, respectively.

- Prior and starting value for R_i (diagonal element of R): The prior for R_i is assumed to be the inverse Gamma distribution:

$$R_i \sim iG(V_R, \tau_R)$$

where $V_R = 0.01$ and $\tau_R = 5$. The starting value of R_i is set equal to one.

Parameters in the RW process of $\ln(h_{i,t})$ (Step 1d)

- Prior and starting value for $\{h_{i,t}\}_{t=1}^{T}$: The prior for the initial value is

$$\ln(h_{i,0}) \sim N(\ln(\mu_{i,0}), 10)$$

  where $\mu_{i,0}$ is i-th diagonal element of $\widehat{v}^{OLS}$. $\mu_{i,0}$ is also used as the starting value for $\{h_{i,t}\}_{t=1}^{T}$.
- Prior and starting value for g_i: The prior for g_i is

$$g_i \sim iG(10^{-4}, 1)$$

  The starting value for g_i is set equal to one.

1.6.4 Simple exercise using VAR (constant-parameter VAR)

As a simple study, I check monetary policy effects by using 7-variable VAR model with fixed coefficients. The data series used in this exercise span 1980:01-1999:12. As for software, I use Eviews throughout this exercise.

The VAR model can be written as follows:

$$X_t = c + A_1 X_{t-1} + ... + A_p X_{t-p} + u_t$$

$X_t = (X_{1,t}, X_{2,t}, ..., X_{N,t})'$ is a set of time series of N observed variables ($N = 7$). u_t is an N-dimensional Gaussian noise with $E[u_t] = 0$, $E[u_t u_t'] = \Sigma$ and $E[u_t u_s'] = 0$ (for t$\neq$s), c is an $N \times 1$ vector of constant term, and A_1, A_2,...,A_p are $N \times N$ autoregressive matrices. The lag length p is set to be 13, following Banbura et al. (2007) (a standard recommendation in the literature).

The 7 variables contained in X_t are listed in Table 1.7. The overnight call rate is used as a monetary policy instrument, and a monetary policy shock is identified by standard recursive identification scheme. As for the transformation of data series, I follow common works in the literature:

- For all variables except for the call rate, the first difference of the log of the seasonally-adjusted time series is taken.
- No transformation is applied to the call rate.

Table 1.7: 7 variables contained in the VAR model

Variables
(1) IIP (Index of Industrial Production)
(2) Employment index
(3) Aggregate CPI
(4) The number of new housing construction started
(5) Overnight call rate
(6) M2
(7) Exchange rate (JPY per USD)

Result for the impulse responses With the setup described in the above, I estimate the VAR model by unrestricted OLS regression. Fig. 1.16 displays the impulse response functions (IRFs) of any element of X_t to a tightening monetary policy shock, where the monetary policy shock is defined as an increase of the call rate by one standard deviation. IIP, the number of housing construction and M2 increases after the shock, and these behaviors do not match the intuition.

Revisiting the treatment of the call rate Based on the unreasonable result shown in the previous subsection, I revisit the treatment of the call rate. As already mentioned, no transformation has been applied to the call rate while the other variables are de-trended. As a de-trending treatment to the call rate, I consider the following options:

(a) Removing a linear trend

(b) Using the Baxter and King bandpass filter[28].

The obtained IRFs in the case (a) are shown in Fig. 1.17. In this case, M2 is found to decrease, which is a reasonable result. However, IIP and the number of housing construction still tend to respond positively. In the case (b), the response behavior becomes much more reasonable, as is shown in Fig. 1.18. The features of the IRFs are summarized as follows:

- IIP and the number of housing construction decrease.
- Employment index, CPI and exchange rate decrease.
- M2 tends to decrease.

[28]The lead-lag length of the filter is set to be 12 months(=1 year), and the bandpass range is to be 6 to 96 months (0.5-8 years).

This result implies that the estimated IRF matches our intuition when the filtered time-series of call rate is used. Based on this, I apply the Baxter-King filter to the monetary policy instrument R_t in the main analysis.

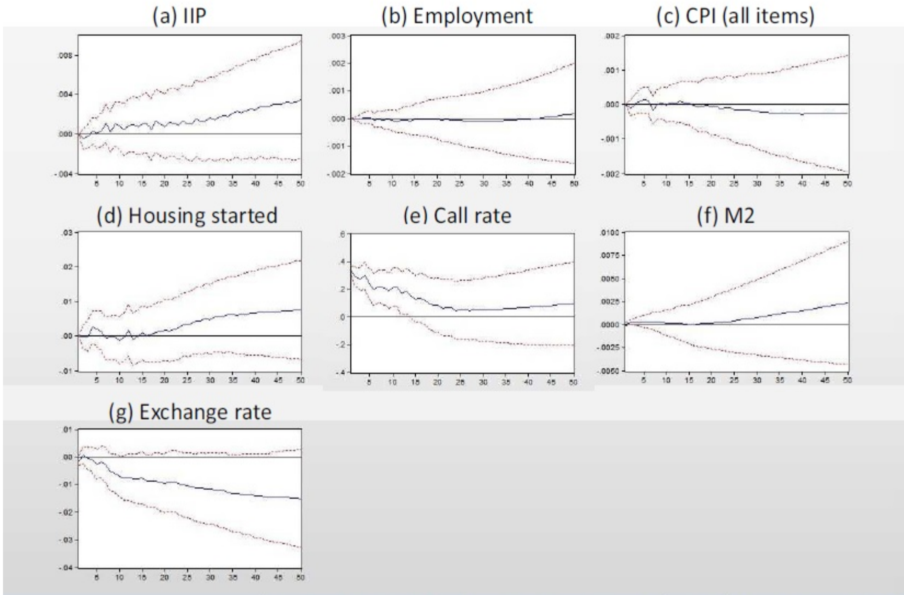


Figure 1.16: Impulse response of the 7 variables obtained from the VAR: The solid line is the point estimate, and the dashed lines are the error bands of two standard deviations.

1.6.5 All other results in the simulation exercise in 1.4.2

When the parameters in IS curve are changed Fig. 1.19 displays the impulse responses when σ is changed (σ: 25.3$\rightarrow$ 9.9). σ^{-1} represents a sensitivity of the output gap to monetary policy. Due to an increase in σ^{-1} (decrease in σ), the response of output gap becomes stronger. This also increases the response strength of inflation via the third term of the Phillips curve.

When the parameters in Phillips curve are changed Fig. 1.20 shows the result when α is changed (α: 0.67$\rightarrow$0.46). Owing to a decrease in inflation's persistence, the speed of inflation reverting to zero becomes more rapid. On the other hand, the responses of output gap and interest rate are hardly affected by the change in α. Fig. 1.21 shows the responses

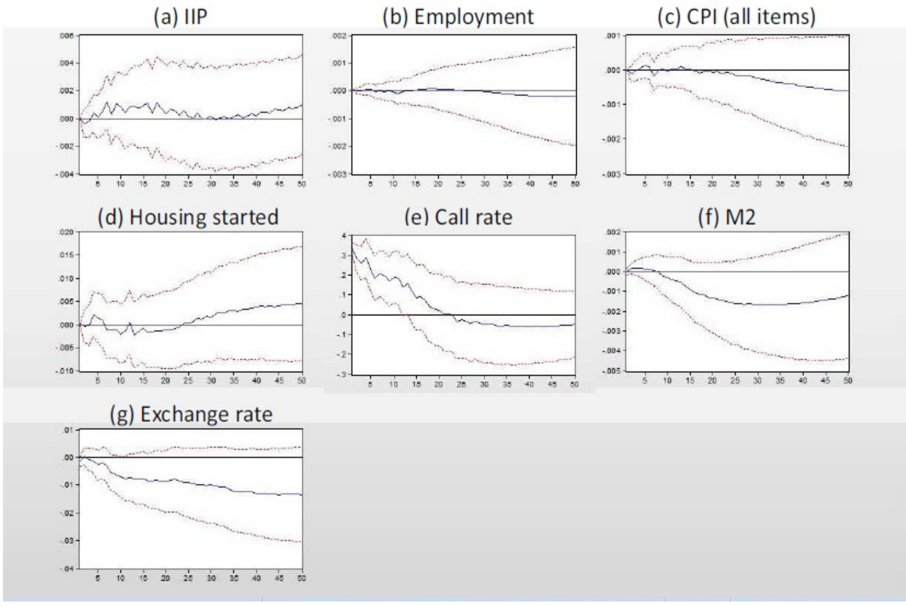


Figure 1.17: Impulse response of the 7 variables obtained from the VAR with linear trend of the call rate removed: The solid line is the point estimate, and the dashed lines are the error bands of two standard deviations.

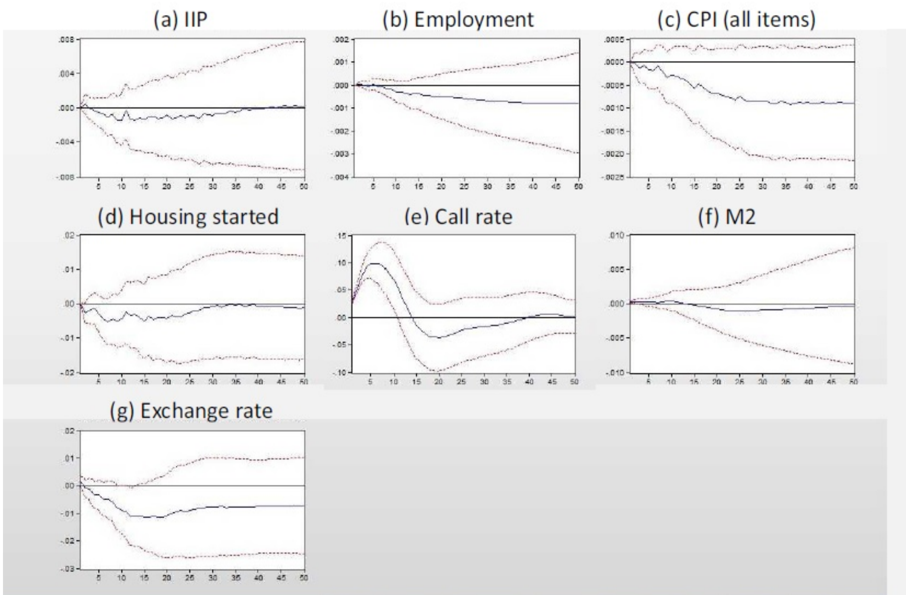
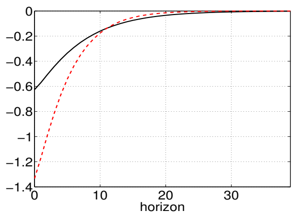


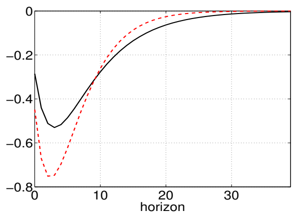
Figure 1.18: Impulse response of the 7 variables obtained from the VAR with the Baxter-King bandpass filter applied to the call rate: The solid line is the point estimate, and the dashed lines are the error bands of two standard deviations.

when κ is changed (κ: 0.037→0.052). κ is a sensitivity of inflation to output gap. The increase in this parameter makes the response of inflation stronger, while those of output and interest rate exhibit little change.

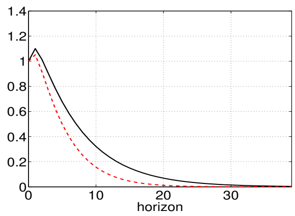
When the parameters in the interest rate rule are changed The responses when ρ is changed (ρ: 0.94→0.87) are shown in Fig. 1.22. ρ represents a persistence of interest rate. The decrease in ρ makes the reverting speed of interest rate to zero becomes more rapid, and furthermore, this reduces the response strength of output gap and inflation. Figs. 1.23 and 1.24 display the impulse responses when ϕ_π and ϕ_y are changed, respectively (ϕ_π: 0.99→1.24, ϕ_y: 1.42→0.79). Fig. 1.23 indicates that monetary policy transmission is hardly affected even when the monetary policy becomes more sensitive to inflation. Fig. 1.24 implies that when the monetary policy is less sensitive to output, the response of inflation becomes stronger while that of output is hardly affected. Fig. 1.25 shows the responses when ρ_R is changed (ρ_R: 0.22→0.09). ρ_R represents a persistence of monetary policy shock. The figure shows that the decrease in ρ_R have no strong effect on monetary policy transmission.



(a) output gap y

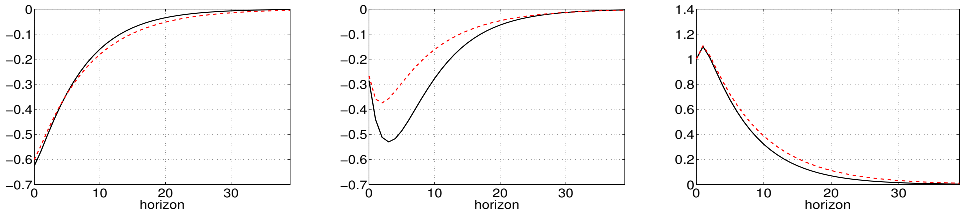


(b) inflation π



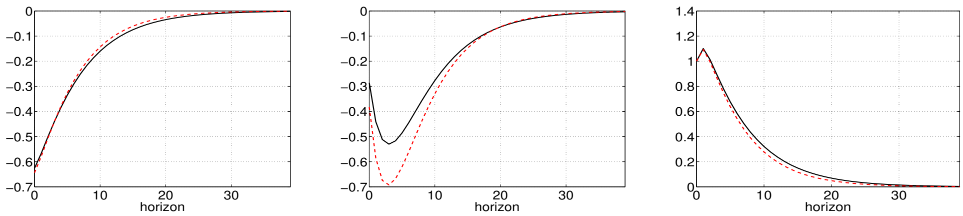
(c) interest rate i_t

Figure 1.19: Impulse responses to a tightening monetary policy shock of 100 basis points when σ is changed (σ: 25.3→ 9.9): Black solid line and red dashed line represent the responses before and after changing σ, respectively.



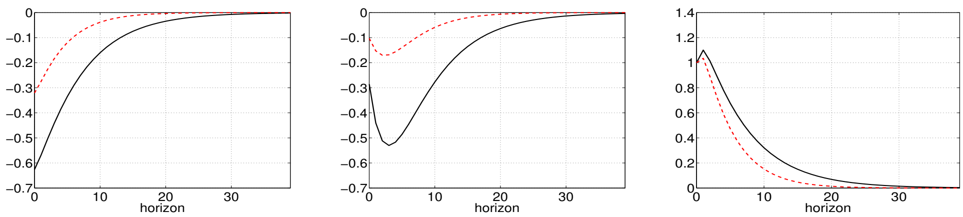
(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.20: Impulse responses to a tightening monetary policy shock of 100 basis points when α is changed (α: 0.67$\rightarrow$ 0.46): Black solid line and red dashed line represent the responses before and after changing α, respectively.



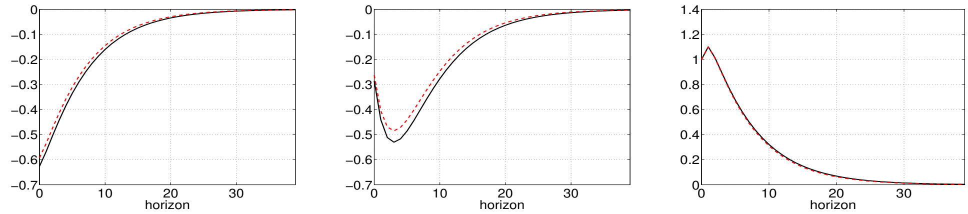
(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.21: Impulse responses to a tightening monetary policy shock of 100 basis points when κ is changed (κ: 0.037$\rightarrow$ 0.052): Black solid line and red dashed line represent the responses before and after changing κ, respectively.



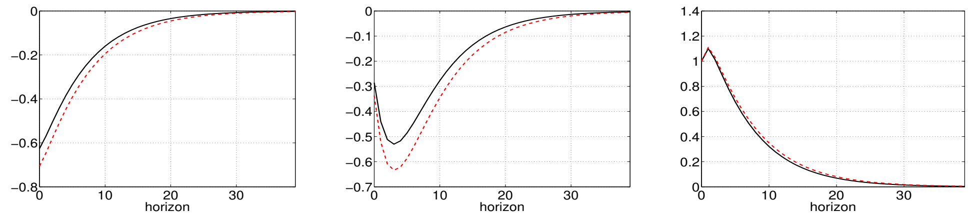
(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.22: Impulse responses to a tightening monetary policy shock of 100 basis points when ρ is changed (ρ: 0.94$\rightarrow$ 0.87): Black solid line and red dashed line represent the responses before and after changing ρ, respectively.



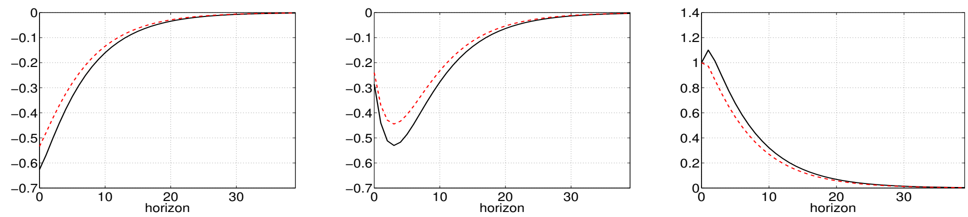
(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.23: Impulse responses to a tightening monetary policy shock of 100 basis points when ϕ_π is changed (ϕ_π: 0.99$\rightarrow$ 1.24): Black solid line and red dashed line represent the responses before and after changing ϕ_π, respectively.



(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.24: Impulse responses to a tightening monetary policy shock of 100 basis points when ϕ_y is changed (ϕ_y: 1.42$\rightarrow$ 0.79): Black solid line and red dashed line represent the responses before and after changing ϕ_y, respectively.



(a) output gap y (b) inflation π (c) interest rate i_t

Figure 1.25: Impulse responses to a tightening monetary policy shock of 100 basis points when ρ_R is changed (ρ_R: 0.22$\rightarrow$ 0.09): Black solid line and red dashed line represent the responses before and after changing ρ_R, respectively.

References for Chapter 1

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70:191–221.

Banbura, M., Giannone, D., and Reichlin, L. (2007). Bayesian VARs with large panels. CEPR Discussion Papers 6326, C.E.P.R. Discussion Papers.

Baumeister, C., Liu, P., and Mumtaz, H. (2013). Changes in the effects of monetary policy on disaggregate price dynamics. *Journal of Economic Dynamics and Control*, 37:543–560.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *Quarterly Journal of Economics*, 123:1005–1060.

Benati, L. and Surico, P. (2009). VAR analysis and the Great Moderation. *American Economic Review*, 99:1636–1652.

Bernanke, B. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50:525–546.

Bernanke, B., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A Factor-augmented Vector Autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120:387–422.

Black, F. (1995). Interest rates as options. *Journal of Finance*, 50:1371–1376.

Blanchard, O. and Simon, J. (2001). The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity*, 32:135–174.

Boivin, J. and Giannoni, M. (2006). Has monetary policy become more effective. *The Review of Economics and Statistics*, 88:54–81.

Canova, F. and Pina, J. (2005). What VAR tell us about DSGE models. *In Diebolt, C., Kyrtsou, C. (Eds.), New Trends in Macroeconomics, Springer, Verlag.*

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–553.

Cogley, T., Primiceri, G., and Sargent, T. (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2:43–69.

Creal, D. and Wu, J. (2015). Estimation of affine term structure models with spanned or unspanned stochastic volatility. *Journal of Econometrics*, 185:60–81.

Duffee, G. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57:405–443.

Ellis, C., Mumtaz, H., and Zabczyk, P. (2014). What lies beneath? A time-varying FAVAR model for the UK transmission mechanism. *Economic Journal*, 0:668–699.

Graeve, F. D. (2017). A structural evaluation of VARs with time-varying parameters and stochastic volatility. Work in progress.

Hamilton, J. and Wu, J. (2014). Testable implications of affine term structure models. *Journal of Econometrics*, 178:231–242.

Jacquier, E., Polson, N., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–417.

Joslin, S., Singleton, K., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. *Review of Financial Studies*, 24:926–970.

Kim, C.-J. and Nelson, C. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press, Cambridge.

Koop, G., Pesaran, H., and Potter, S. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74:119–147.

Lubik, T. and Schorfheide, F. (2004). Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review*, 94:190–217.

Perez-Quiros, G. and McConnell, M. M. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review*, 90:1464–1476.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72:821–852.

Sims, C. A. (2002). Solving linear rational expectations models. *Computational Economics*, 20:1–20.

Stock, J. and Watson, M. (2005). An empirical comparison of methods for forecasting using many predictors. Manuscript, Princeton University.

Svensson, L. (1995). Estimating forward interest rates with the extended Nelson and Siegel method. *Sveriges Riksbank Quarterly Review*, 3:13–26.

Wright, J. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review*, 101:1514–1534.

Wu, J. and Xia, F. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit, and Banking*, 48.

Yamamura, T. (2018). Evaluating FAVAR with time-varying parameters and stochastic volatility. Working Paper.

Chapter 2

Evaluating FAVAR with Time-Varying Parameters and Stochastic Volatility

2.1 Introduction

In macroeconomics, VAR (Vector Autoregression) and FAVAR (Factor Augmented Vector Autoregression) models have been used in a wide range of empirical analyses. Examples of their application include the transmission of monetary policy, fiscal policy, and oil prices in the real economy. However, in such analyses, it needs to be considered carefully as to whether the dynamics behind such economic phenomena are time-invariant. If they vary over time, the empirical models should be extended so as to take this variation into account. The literature has attempted to extend the models in several ways. One popular approach is the models allowing for time-varying parameters (TVP) and stochastic volatility (SV): TVP/SV-VAR and TVP/SV-FAVAR. Incorporating time variation makes the number of unknown parameters incomparably larger, but the models can be estimated by a standard Bayesian approach without imposing harmful assumptions. Using this methodology, many studies have approached several important economic topics such as the Great Moderation (Cogley and Sargent (2005), Primiceri (2005), Gali and Gambetti (2009), Bianchi et al. (2009), Canova and Gambetti (2010), Baumeister et al. (2013), Ellis et al. (2014)); quantitative easing (Kapetanios et al. (2012), Baumeister and Benati (2013)); oil prices (Baumeister and Peersman (2013)); and fiscal policy (Kilem et al. (2015)).

Although both TVP/SV-VAR and TVP/SV-FAVAR have been used in many empirical research studies, some researchers have noticed that the time variations identified by these empirical models should be interpreted carefully. Benati and Surico (2009) conduct a simulation-based experiment, where they investigate the nature of VAR results when there is a shift from passive to active monetary policy regimes in the data-generating process (DGP)[1]. Although monetary policy effects on the economy change at the underlying level, the estimated impulse response functions (IRFs) do not identify it. This result is related to a structural difference between the structural model (the underlying DSGE) and its reduced-form VAR representation. As they point out, a change in the systematic component of monetary policy does not necessarily bear a clear-cut relation to the one in the VAR coefficients. As a consequence, time variations identified by the empirical model are not always associated with the ones in the structural model in a direct manner. Due to this, the estimated VAR can provide a misleading analysis result, as revealed by the experiment of Benati and Surico (2009).

Regarding this argument, Canova (2006) insists that the choice of their experimental setup is not the most relevant one[2]. As one reason, he points out that the DGP used by Benati and Surico (2009) includes indeterminate equilibrium which is associated with the passive policy regime. As argued by Graeve (2017), reliance on sunspots in one regime complicates representation issues in terms of both DSGE and VAR. Therefore, the experimental results of Benati and Surico (2009) may not necessarily bring about a general implication on the nature of VAR analysis results. Using this background, Graeve (2017) conducts a simulation-based experiment with a more general setup. He examines the performance of TVP/SV-VAR against several types of time variation at the DGP level[3],

[1]For this investigation, they first estimate the New Keynesian (NK) model with U.S. data for each of the pre- and post-1980s periods. Then, by using the estimated NK model, they prepare a simulation sample, with which the VAR results are examined.

[2]In this unpublished note, he provides comments and criticisms to the Benati and Surico's work.

[3]This is the first work to examine TVP/SV-VAR from a theoretical perspective. Note that Benati and Surico (2009) use a constant-parameter VAR throughout their experiment.

and finds that TVP/SV-VAR can usually capture time variation in the underlying dynamics in an adequate manner.

As such, Graeve (2017) provides an evidence in favor of the time-varying VAR framework. However, there are at least two remaining issues which have not been addressed in the literature. First, Graeve (2017) uses small-scale DSGEs as the DGP[4]. When a scale of the underlying economy is small, VAR model tends to have information set enough to capture a structural shock of interest (ex. monetary policy shock). However, it is not necessarily the case when the economy is large in its scale, and yet, he has not considered this possibility. The second question regards the implications of using TVP/SV-FAVAR instead. By virtue of factor-augmented structure, TVP/SV-FAVAR is allowed to include large information sets, and due to this feature, it may be more suitable than TVP/SV-VAR for identifying a structural shock of interest. It would be worthwhile to look into an advantage of using TVP/SV-FAVAR from this perspective. However, no studies in the literature have examined it.

Against this backdrop, this chapter's study is devoted to investigating the performance of TVP/SV-FAVAR in comparison to that of TVP/SV-VAR. As in Benati and Surico (2009), the performance is evaluated in terms of the ability to capture the time variation in monetary policy effects, but note that indeterminate equilibrium is never considered throughout this study. The analysis is conducted mainly through Monte Carlo (MC)-based experiments using open-economy DSGE as the DGP. Since open-economy DSGE is a medium-scale model, this allows me to mimic a situation with the following features: (i) the true economy contains many variables; (ii) VAR does not include all of them; and (iii) FAVAR is, on the contrary, allowed to do so. As the main result, the experiments reveal that TVP/SV-FAVAR adequately detects the time variation in monetary policy effects, whereas TVP/SV-VAR does not. Subsequently, this result is further interpreted from the perspective of the information sufficiency of the two models.

This study brings a valuable contribution to empirical studies focusing on the time variations in any sort of economic phenomena. In particular, it is of a great relevance to the literature of the Great Moderation. The Great Moderation is a marked decline in the volatility of macroeconomic variables (such as output and inflation) during the early 1980s and the

[4]He uses the New Keynesian (NK) and Real Business Cycle (RBC) models.

subsequent periods in the U.S.[5] As for the source of this phenomenon, many earlier works using VAR methods provide *the good luck hypothesis* (see Stock and Watson (2002), Primiceri (2005), Sims and Zha (2006), and Gambetti et al. (2008)), which attributes the macroeconomic stability to an exogenous cause (i.e. a decline in the volatility of exogenous shocks). However, more recent works using TVP/SV-FAVAR provide a different analysis result. One pioneering work with this empirical framework is Baumeister et al. (2013). They examine the time variations in monetary policy effects from the 1970s to the 2000s in the U.S., and observe a significant variation during the 1980s[6]. Regarding why VAR and FAVAR provide such inconsistent results, it has not been fully discussed in the literature. However, as indicated by my MC experiments, the performance of VAR may be problematic, and this would be a possible explanation for the above inconsistency.

In order to reinforce the findings of my MC experiments, I also conduct the empirical exercise with an application to Japan's data from the 1970s to the 2000s, as an example exhibiting different performances of TVP/SV-FAVAR and TVP/SV-VAR[7]. The IRF analysis reveals that in the case of TVP/SV-FAVAR, the response of the aggregate price to monetary policy shock significantly strengthens during the early 1990s, and the timing of this change is associated with the bubble collapse, which happened in 1990-91 in Japan. On the other hand, TVP/SV-VAR indicates no evidence of time variations in monetary policy transmission. This example seems to reproduce the findings in the empirical literature of the Great Moderation, in the sense that the FAVAR result suggests a significant variation in the monetary policy effects whereas the VAR does not necessarily identify it.

This chapter is organized as follows. In Section 2.2, the empirical models (TVP/SV-FAVAR and TVP/SV-VAR) are reviewed. In Section 2.3, MC-based experiments are conducted, and TVP/SV-FAVAR displays a good performance whereas it is not the case for TVP/SV-VAR.

[5] Many papers report this phenomenon. See, for example, Perez-Quiros and McConnell (2000) and Blanchard and Simon (2001).

[6] Ellis et al. (2014) and Yamamura (2017) analyze UK and Japan's data with TVP/SV-FAVAR, and they also report a significant variation in monetary policy propagation during the early 1990s. Note that Yamamura (2017)'s work is reported in Chapter 1 of this thesis.

[7] Note that the analysis with TVP/SV-FAVAR was already conducted in Chapter 1. In this chapter, the analysis with TVP/SV/VAR is newly added.

In Section 2.4, the obtained results are interpreted from the perspective of the information sufficiency of the two empirical models. Section 2.5 gives an example of the empirical application using Japan's data, and Section 2.6 provides a conclusion.

2.2 Empirical models

This section gives an overview of TVP/SV-FAVAR and TVP/SV-VAR.

2.2.1 TVP/SV-FAVAR

The specification of TVP/SV-FAVAR is the same as the one which I used in Chapter 1 (see 1.3.1). First, the observation equation is

$$X_t = \Lambda^f F_t + \Psi^R R_t + e_t \ , \quad e_t \sim N(0, R) \qquad (t = 1, ..., T) \tag{2.1}$$

where $X_t = [X_{1t} \ ... \ X_{Nt}]'$ is a panel of N observed variables, $F_t = [F_t^1 \ ... \ F_t^K]'$ denote K latent factors, and R_t is interest rate. The disturbances $e_t = [e_{1t} \ ... \ e_{Nt}]'$ are *i.i.d.* with $E[e_t] = 0$ and $E[e_t e_t'] = R$, where the $N \times N$ matrix R is assumed to be diagonal. For unique identification of latent factors, it is assumed that the upper $K \times K$ block of Λ^f is identity matrix, and also that the upper $K \times 1$ block of Ψ^R is zero[8]. The transition equation is given by

$$Z_t = c_t + B_{1,t} Z_{t-1} + ... + B_{L,t} Z_{t-L} + v_t \ , \quad v_t \sim N(0, \Omega_t) \tag{2.2}$$

where Z_t denotes the common factors made up of latent factors F_t and interest rate R_t. Note that this equation allows for time-varying parameters (c_t, $B_{k,t}$ with $k = 1, ..., L$) and stochastic volatility (Ω_t). As a common choice in the literature, the lag length is set equal to two ($L = 2$)[9]. The volatility matrix Ω_t is factored as

$$\Omega_t = A_t^{-1} H_t (A_t^{-1})' \tag{2.3}$$

[8]For detail, see Bernanke et al. (2005).

[9]This choice is mainly motivated from a computational perspective. For more detail, see Primiceri (2005), Baumeister et al. (2013) and Ellis et al. (2014).

where H_t is diagonal and A_t is lower-triangular:

$$H_t = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t} \end{bmatrix} \quad , \quad A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1,t} & \cdots & a_{nn-1,t} & 1 \end{bmatrix} \tag{2.4}$$

Following Primiceri (2005), Baumeister et al. (2013) and Ellis et al. (2014), the time evolution of the VAR coefficients and volatility matrix is specified as

$$\begin{aligned} \beta_t &= \beta_{t-1} + \eta_t \\ \alpha_t &= \alpha_{t-1} + \tau_t \\ \log \sigma_t &= \log \sigma_{t-1} + \varepsilon_t \end{aligned} \tag{2.5}$$

where β_t stacks all of the VAR coefficients, α_t stacks non-zero and non-one elements of the matrix A_t, and σ_t stacks diagonal components of the matrix H_t. All of the innovations in the model are assumed to be jointly normally distributed, and their variance covariance matrices are specified by

$$V \equiv Var\left(\begin{bmatrix} u_t \\ \eta_t \\ \tau_t \\ \varepsilon_t \end{bmatrix} \right) = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & D & 0 \\ 0 & 0 & 0 & G \end{bmatrix} \tag{2.6}$$

where G is assumed to be diagonal ($G = \text{diag}(g_i)$).

2.2.2 TVP/SV-VAR

As for TVP/SV-VAR, the specification follows Primiceri (2005). The equation system is given by

$$X_t = c_t + B_{1,t}X_{t-1} + ... + B_{L,t}X_{t-L} + v_t \quad , \quad v_t \sim N(0, \Omega_t) \tag{2.7}$$

where the vector X_t consists of the observed variables. Regarding time variation in the coefficients and shock variance, the same assumptions are adopted as in the transition equation in TVP/SV-FAVAR as summarized by Eqs. (2.3)-(2.6). As in TVP/SV-FAVAR, the lag length is set equal to two ($L = 2$).

Table 2.1: Endogenous variables in the open-economy DSGE.

Notation	Definition
C_t	consumption
Y_t	output
π_t	inflation
$\pi_{H,t}$	inflation for domestic goods
$\pi_{F,t}$	inflation for imported goods
S_t	terms of trade
$\Psi_{F,t}$	price gap
$\widetilde{q}_t$	real exchange rate
R_t	nominal interest rate
Y_t^*	output in foreign economy
π_t^*	inflation in foreign economy
R_t^*	nominal interest rate in foreign economy

2.3 Monte Carlo (MC)-based experiments

In this section, MC-based experiments are conducted to examine whether TVP/SV-FAVAR and TVP/SV-VAR correctly detect time variation in monetary policy transmission.

2.3.1 Open-economy DSGE model

As the data-generation process (DGP), I use the open-economy DSGE model developed by Justiniano and Preston (2010). This model is a generalization of the model of Monacelli (2005), where the generalization includes the introduction of incomplete asset markets, habit formation, and the indexation of prices to past inflation. The model consists of eight exogenous shocks and twelve macroeconomic endogenous variables. All of the endogenous variables are listed in Table 2.1. The model is calibrated by the estimates of Justiniano and Preston (2010), but as described below, a certain form of time variation is additionally imposed in the monetary policy rule.

In the open-economy DSGE, the monetary policy rule is governed by

$$\frac{R_t}{\overline{R}} = \left(\frac{R_{t-1}}{\overline{R}}\right)^{\rho_i} \cdot \left(\frac{1+\pi_t}{1+\overline{\pi}}\right)^{\psi_\pi} \cdot \left(\frac{Y_t}{\overline{Y}}\right)^{\psi_y} \cdot \left(\frac{Y_t}{Y_{t-1}}\right)^{\psi_{\Delta y}} \cdot \left(\frac{\widetilde{e}_t}{\widetilde{e}_{t-1}}\right)^{\psi_{\Delta e}} \times \widetilde{\varepsilon}_{M,t} \tag{2.8}$$

where R_t, Y_t, π_t, $\widetilde{e}_t$, and $\widetilde{\varepsilon}_{M,t}$ represent interest rate, output, inflation, nominal exchange rate, and monetary policy shock, respectively[10]. Note also that $\overline{R}$, $\overline{Y}$ and $\overline{\pi}$ indicate steady-state level. As per time variation in the policy rule, the following structural break is introduced in the parameter ψ_π:

$$\psi_\pi = \begin{cases} 1.8 & \text{(for } 1 \leq t \leq 125; \quad \text{Regime 1)} \\ 0.45 & \text{(for } 126 \leq t \leq 250; \text{ Regime 2)} \end{cases} \tag{2.9}$$

The value in Regime 1 ($\psi_\pi = 1.8$) is the estimate of Justiniano and Preston (2010). In Regime 2, ψ_π is set four times smaller than that ($\psi_\pi = 0.45$). Note that the choice of *four times change* is based on the observation of Fernandez-Villaverde and Rubio-Ramirez (2008)[11]. Fig. 2.1 shows the theoretical IRFs of the three major variables (output, inflation, and interest rate) to monetary policy shock. Throughout this study, monetary policy shock is defined by an increase of 100 basis points in the interest rate. It exhibits that the variation in the parameter ψ_π significantly strengthens the impulse responses of output and inflation.

The model is solved separately by the regimes[12]. Then, by using the solution form, the model is simulated 100 times, by which 100 pseudo-data samples (simulation samples) are prepared. In each simulation, 500 observations are generated for each regime, using the exogenous shocks drawn from normal distribution. The first 375 observations are discarded to reduce the impact of initial conditions. By this procedure, the sample size is set equal to 250 ($T = 250$).

[10] $\widetilde{e}_t$ is defined by $\widetilde{e}_t = \widetilde{q}_t P_t / P_t^*$, where $\widetilde{q}_t$ is real exchange rate, and P_t and P_t^* denote the aggregate price in domestic and foreign economies, respectively.

[11] They observed that γ_π (equivalent parameter to ψ_π) changes by 4-5 times through the Great Moderation.

[12] The solution form of the model is obtained by converting the equation system of the DGP to the following VAR form:

$$\widetilde{X}_t = A\widetilde{X}_{t-1} + Bu_t$$

where $\widetilde{X}_t$ and u_t are vectors of endogenous variables and structural shocks, respectively. This solution form is often called the “Sims’ canonical form”. As for how to obtain the matrices A and B, see Sims (2002).

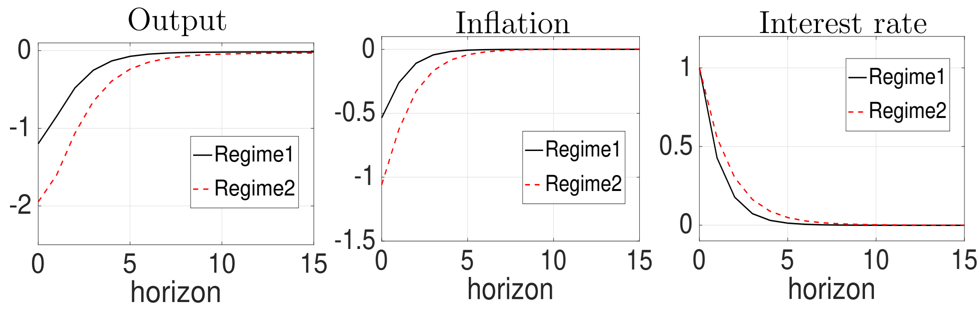


Figure 2.1: Theoretical impulse response functions (IRFs) of output, inflation, and interest rate to monetary policy shock in the open-economy DSGE. The black solid line and the red dashed line represent the responses in Regimes 1 and 2, respectively. Throughout this study, the monetary policy shock is defined by an increase of 100 basis points in the interest rate.

2.3.2 Estimation of TVP/SV-FAVAR

Identification scheme First, the observed vector X_t and the common factors Z_t are defined by

$$X_t = \{Y_t^*, \pi_t^*, R_t^*, C_t, Y_t, \pi_t, \pi_{H,t}, \pi_{F,t}, S_t, \Psi_{F,t}, \widetilde{q}_t\}' \tag{2.10}$$

$$Z_t = \{R_t, F_t\}' = \{R_t, F_t^1, ..., F_t^K\}' \tag{2.11}$$

For the definition of each variable, see Table 2.1. Note also that log difference transformation is applied to all of the variables except for the interest rate R_t[13]. Then, to identify monetary policy shock, the sign restrictions scheme is adopted. Following Canova and Nicol (2002), Uhlig (2005), and Ellis et al. (2014), restrictions are placed on the contemporaneous response of the observed variables. As restrictions, it is assumed that a tightening monetary policy shock increases the interest rate, and decreases the output and inflation on impact:

$$\begin{aligned} (IRF)^Y_{t,h=0} &< 0 \\ (IRF)^\pi_{t,h=0} &< 0 \\ (IRF)^R_{t,h=0} &> 0 \end{aligned} \tag{2.12}$$

where $(IRF)^s_{t,h=0}$ represents the impulse response of the variable s ($= Y, \pi,$ or R) to monetary policy shock at time t for horizon $h = 0$ (i.e.

[13]In the case of using a large dataset (such as FAVAR), this transformation is recommended to induce stationarity (see Baumeister et al. (2013) and Ellis et al. (2014)).

contemporaneous response). Note that these restrictions are in line with the true properties of the DGP.

To impose the above restrictions, I undertake the procedure used by Ellis et al. (2014). The restrictions are introduced as follows. First, the Cholesky decomposition is applied to the VAR covariance matrix Ω_t:

$$\Omega_t = P_t P_t'$$

Then, a $n \times n$ matrix ($\equiv J$) is drawn from the $N(0,1)$ distribution. Next, the QR decomposition is applied to J (i.e. $J = QR$ with $QQ' = I$), by which a candidate covariance matrix is decomposed as $\Omega_t = \widetilde{P}_t \widetilde{P}_t'$ with $\widetilde{P}_t = P_t Q$. Using this candidate matrix, the contemporaneous impulse responses of the observed variables to monetary policy shock are calculated. If these responses satisfy the restrictions (2.12), $\widetilde{P}_t$ is stored. This procedure is repeated until 100 $\widetilde{P}_t$ matrices are stored. Finally, the matrix $\widetilde{P}_t$ is defined by elements closest to the median across the 100 stored matrices.

Estimation procedure The estimation procedure is similar to the one I used in Chapter 1. To begin with, the factor loadings $\Gamma \left(= (\Psi^R, \Lambda^f)\right)$ and the variance R ($= \text{diag}(R_i)$) are sampled conditionally on given latent factors. Γ and R_i are drawn from the normal and inverse gamma distributions respectively. The VAR coefficients (β_t), the off-diagonal elements of A_t ($a_{ij,t}$), and the covariance matrices of disturbances in their random-walk process (Q and D) are subsequently drawn through the method developed by Carter and Kohn (1994), while $h_{i,t}$ and g_i are simulated using the scheme described in Jacquier et al. (1994). Using the procedure described in the previous paragraph, the variance matrix Ω_t is decomposed as $\Omega_t = \widetilde{P}_t \widetilde{P}_t'$. The latent factors F_t are sampled by the algorithm of Carter and Kohn (1994). The above steps are iterated 10,000 times, with the first 9,000 draws removed as a burn-in (M =10,000 and M_0=9,000). Note that more details of the algorithm are described in 1.6.2 and 1.6.3.

As for the number of latent factors K, it is set equal to two ($K = 2$). This choice is based on the exercises in 2.7.2, where I examine TVP/SV-FAVAR in the three cases of K ($K = 1, 2, 3$)[14]. The exercises indicate

[14]When $K \geq 4$, the computation time explodes and the model estimation becomes

that the model adequately detects time variation in monetary policy transmissions if $K \geq 2$. However, another finding is that when $K = 3$, the estimation uncertainty is clearly larger than in the case of $K = 2$. Based on these findings, I choose $K = 2$ as the case yielding TVP/SV-FAVAR's best performance.

2.3.3 Estimation of TVP/SV-VAR

In TVP/SV-VAR, the VAR vector X_t is defined by

$$X_t = \{R_t, Y_t, \pi_t\}' \tag{2.13}$$

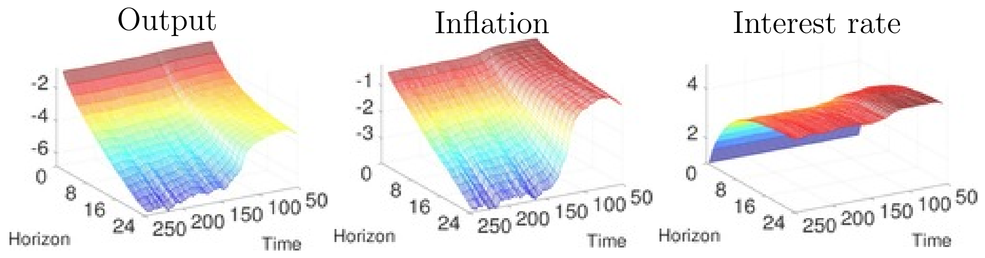
As in TVP/SV-FAVAR, monetary policy shock is identified by the sign restrictions scheme. As for estimation algorithm, it is exactly the same as the one for the transition equation part in TVP/SV-FAVAR.

2.3.4 Results

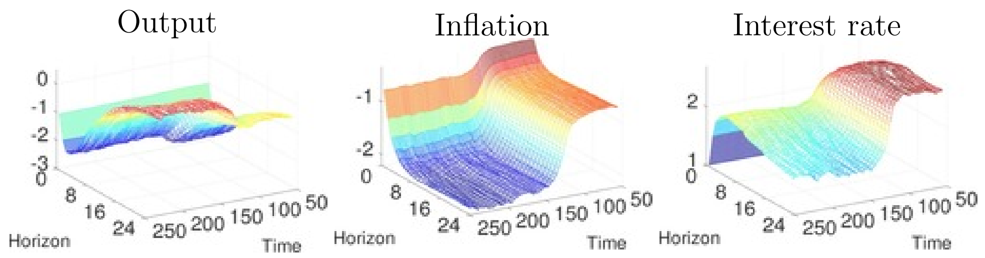
Using the 100 simulation samples, I perform 100 pseudo-experiments where, in each experiment, TVP/SV-FAVAR and TVP/SV-VAR are estimated using the Bayesian approach. Fig. 2.2(a) displays the estimated median cumulative IRFs to monetary policy shock across the 100 experiments with TVP/SV-FAVAR[15]. As already mentioned in 2.3.1, monetary policy shock is defined by an increase of 100 basis points in the interest rate. Note also that the first 50 periods ($t = 1-50$) are not included in each plot, because they are used as a training sample. In the long-horizon region, the responses of output and inflation strengthen around $t = 125$, which is consistent in both the timing and direction of variation, with the true change in the DSGE (see Fig. 2.1). Fig. 2.2(b) shows the corresponding plots obtained by TVP/SV-VAR. It suggests that the response of inflation strengthens around $t = 125$, which is the same feature as with TVP/SV-FAVAR. However, regarding output, the shape of its response function is strange. Furthermore, the interest rate also exhibits

unstable.

[15] As for IRFs of output and inflation, some constant correction factors are applied to them. The merit of using this correction is that the long-run cumulative IRF becomes a well-defined measure of total impact of monetary policy shock on each observed variable. Note also that this correction has no impact on the time variation in the estimated cumulative IRFs. For more detail, see 2.7.1.



(a) TVP/SV-FAVAR



(b) TVP/SV-VAR

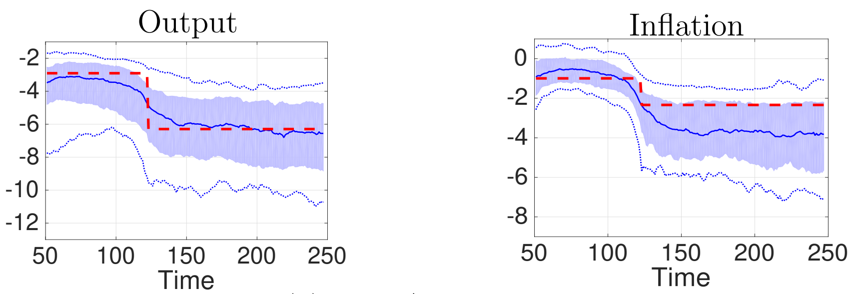
Figure 2.2: Time-varying median cumulative impulse responses to monetary policy shock across the 100 pseudo-experiments using (a) TVP/SV-FAVAR and (b) TVP/SV-VAR.

a sudden change at $t = 125$ in its response. This should be also regarded as strange, because the theoretical IRF of interest rate is not affected significantly by the change in ψ_π (see Fig. 2.1).

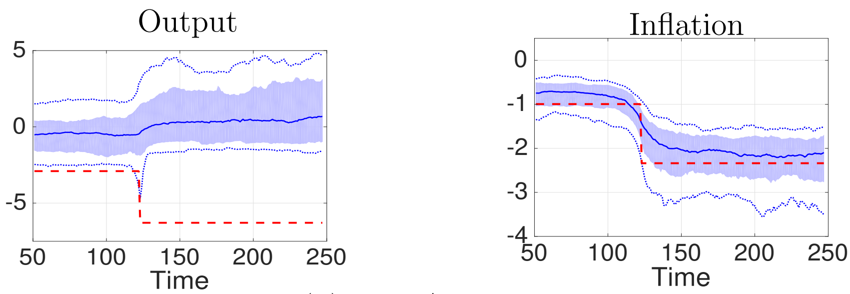
As to the responses of the output and inflation, I also check their time variation at a fixed impulse horizon of $h = 24$, as shown in Fig. 2.3. In this figure, the estimated responses are compared with the true one, as indicated by the red line. TVP/SV-FAVAR exhibits that the estimated responses are consistent with the true one within the bounds of estimation uncertainty, in terms of both output and inflation. However, in the case of TVP/SV-VAR, the estimated result for the output deviates from the true response, and the gap between them cannot be explained by estimation uncertainty.

2.3.5 Robustness check

As a robustness check, two alternative exercises are conducted with the following setup: (i) tightened prior is imposed on time variation process of VAR coefficients β_t, and (ii) smoothed time variation is assumed in



(a) TVP/SV-FAVAR



(b) TVP/SV-VAR

Figure 2.3: Time-varying cumulative impulse responses at the impulse horizon $h = 24$. The responses are obtained by the 100 pseudo-experiments using (a) TVP/SV-FAVAR and (b) TVP/SV-VAR. The blue solid line, shaded area, and dotted lines represent the median, 16/84-percentiles, and 5/95-th percentiles of the estimated responses, respectively. The red dashed line indicates the true response in the DSGE.

the parameter ψ_π.

Tightened prior In TVP/SV-FAVAR and TVP/SV-VAR, the time-varying coefficients β_t play an important role in capturing time variation in monetary policy transmission. As described in Section 2.2, their volatility is controlled by the covariance matrix Q. In the *baseline exercise* conducted in the above subsections, the prior of Q is defined by the inverse-Wishart distribution[16]:

$$Q \sim iW(Q_0, T_0)$$

where $Q_0 = \tau \times Var[\widehat{\beta}^{OLS}] \times T_0$ with $\tau = 3.5 \times 10^{-4}$, and T_0 is a length of the training sample[17]. To perform a robustness check, τ is tightened to 1.0×10^{-4}. Note that tighter prior of Q restricts the time-varying coefficients β_t to be less volatile.

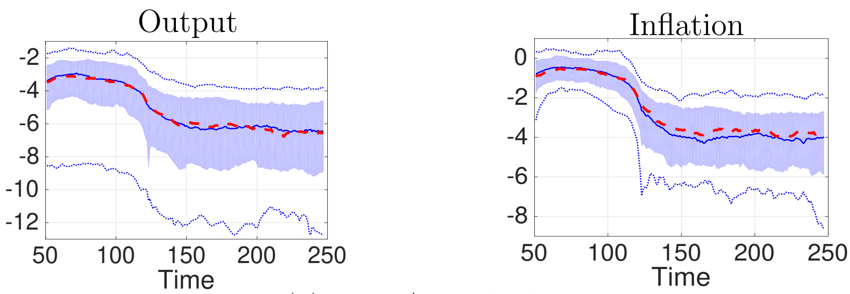
Then, MC exercise is conducted, where the experimental setup is the same as in the baseline exercise except for the prior of Q. Fig. 2.4 shows the obtained cumulative responses at an impulse horizon $h = 24$. In each plot, the median estimated response (blue solid line) is almost the same as that of the baseline exercise (red dashed line). This suggests that the experimental results are hardly affected by the prior setup.

DGP with smoothed time variation In the baseline exercise, a structural change was imposed in the monetary policy rule (see Eq. (2.9)). As another option, let us consider the following smoothed time variation:

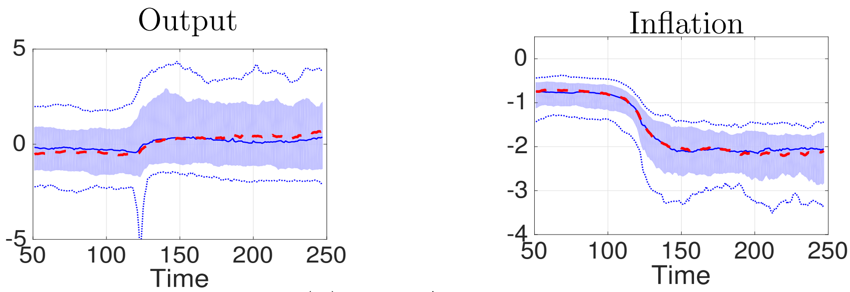
$$\begin{cases} \psi_{\pi,t} &= A + B \cdot \sin\left(\dfrac{2\pi}{T}t\right) \qquad (T = 250) \\ A &= \dfrac{1}{2} \cdot (\psi_{max} + \psi_{min}) \\ B &= \dfrac{1}{2} \cdot (\psi_{max} - \psi_{min}) \end{cases}$$

[16]As mentioned in 2.3.2, this follows a standard recommendation in the literature. For more detail, see 1.6.3.

[17]The periods t = 1-50 are used for the training. Moreover, $\widehat{\beta}^{OLS}$ represents an OLS estimate, which is obtained by the OLS estimation over the training sample via a fixed-coefficient VAR model made up of the interest rate and the principal components (i.e. the VAR vector is defined as $Z_t = \{R_t, (PC)_t\}$.). For more detail, see 1.6.3.



(a) TVP/SV-FAVAR



(b) TVP/SV-VAR

Figure 2.4: Time-varying cumulative impulse responses at the impulse horizon $h = 24$ when the prior for Q is tightened ($\tau = 1.0 \times 10^{-4}$). The responses are obtained by the 100 pseudo-experiments using (a) TVP/SV-FAVAR and (b) TVP/SV-VAR. The blue solid line, shaded area, and dotted lines represent the median, 16/84-percentiles, and 5/95-th percentiles of the estimated responses, respectively. The red dashed line indicates the estimated median response obtained in the baseline MC exercise.

where $\psi_{max} = 1.8$ and $\psi_{min} = 0.45$. With this setup, the model is simulated 100 times[18], by which 100 pseudo-data samples are obtained. In each simulation, 750 observations are generated by using exogenous shocks drawn from the normal distribution, and the first 500 observations are discarded to reduce the impact of the initial conditions. By this procedure, the sample size is set as equal to 250.

Using this DGP, I perform the 100 pseudo-experiments with each of TVP/SV-FAVAR and TVP/SV-VAR. Fig. 2.5 shows the obtained time-varying cumulative IRFs. In the case of TVP/SV-FAVAR, the responses of the output and inflation exhibit a smoothed time variation but that of the interest rate hardly varies. These results are consistent with the true properties of the DGP. In the case of TVP/SV-VAR, the response of the inflation varies smoothly over time in an expected manner. However, the output exhibits an unreasonable feature in terms of the shape of its response function. As another irrelevant result, the response of interest rate is observed to vary over time. Fig. 2.6 illustrates the time variation in the cumulative responses of output and inflation at the impulse horizon $h = 24$. The estimated result with TVP/SV-FAVAR is in agreement with the truth. However, TVP/SV-VAR shows that the estimated response of output is not consistent with the true response, whose inconsistency cannot be explained by estimation uncertainty. In summary, Figs. 2.5 and 2.6 provide the same implication as in the baseline exercise, indicating that TVP/SV-FAVAR is superior to TVP/SV-VAR in terms of identifying time variation in monetary policy effects.

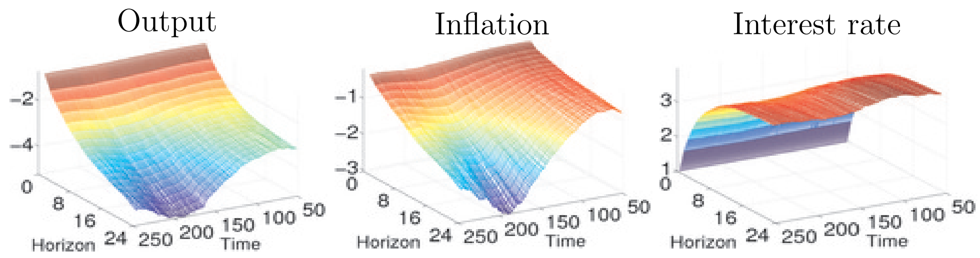
Further exercise As a further robustness check, I also conduct an MC exercise using Smets-Wouters model (2007)[19] as DGP. For details, see 2.7.3.

2.4 Interpretation

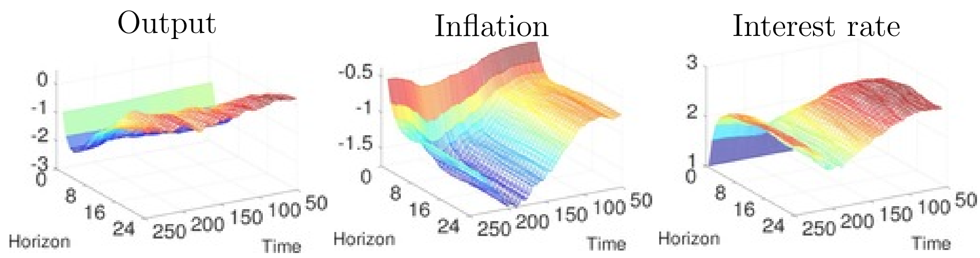
In Section 2.3, TVP/SV-FAVAR displayed a good performance in terms of detecting time variation in monetary policy transmission, whereas TVP/SV-VAR did not. In this section, these results are interpreted

[18]Note that ψ_π takes a different value for each t. Therefore, the solution form of the model is prepared for each t in a separate manner.

[19]For detail of this model, see Smets and Wouters (2007).

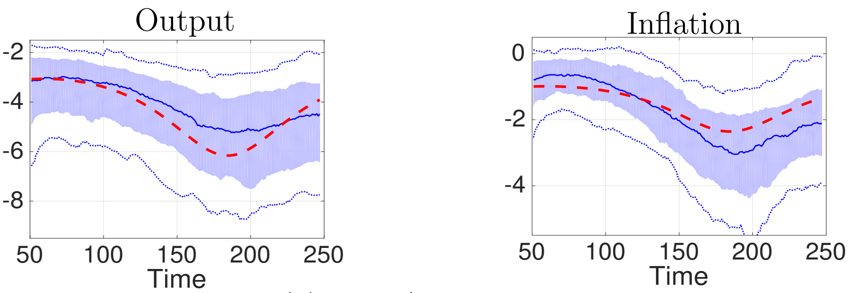


(a) TVP/SV-FAVAR

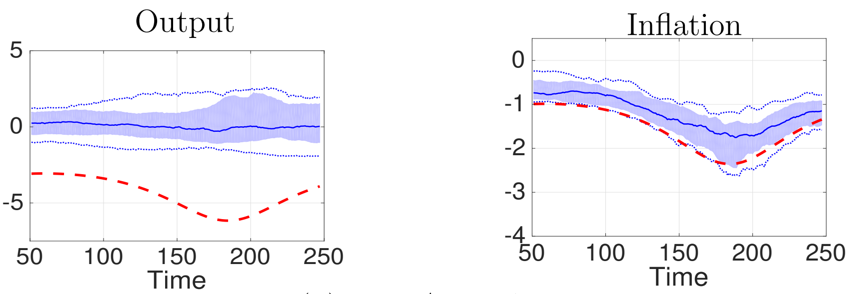


(b) TVP/SV-VAR

Figure 2.5: Time-varying median cumulative impulse responses to monetary policy shock across the 100 pseudo-experiments where smoothed time variation is imposed in ψ_π. The estimated responses from (a) TVP/SV-FAVAR and (b) TVP/SV-VAR are depicted.



(a) TVP/SV-FAVAR



(b) TVP/SV-VAR

Figure 2.6: Time-varying cumulative impulse responses at the impulse horizon $h = 24$ across the 100 pseudo-experiments where smoothed time variation is imposed in ψ_π. The estimated responses from (a) TVP/SV-FAVAR and (b) TVP/SV-VAR are depicted. The blue solid line, shaded area, and dotted lines represent the median, 16/84-percentiles, and 5/95-th percentiles of the estimated responses, respectively. The red dashed line indicates the true response in the DSGE.

from the perspective of information sufficiency of the two models. For this analysis, I use the concept of *informational deficiency* and a technique to quantitatively evaluate it, both of which were recently proposed by Forni et al. (2016). In Subsection 2.4.1, the informational deficiency of the 3-variable VAR is evaluated in terms of identifying monetary policy shock. The same exercise is also applied to the VAR system comprised of the principal components and interest rate, by which the deficiency of the FAVAR model is also examined. Based on the obtained results, Subsection 2.4.2 provides an interpretation of the performance of the two models in the MC experiments.

Regarding the evaluation of the informational deficiency (or sufficiency), the technique presented by Forni et al. (2016) is not the only approach proposed in the literature. One popular method is the Granger causality test developed by Giannone and Reichlin (2006) and Forni and Gambetti (2014). Moreover, Canova and Hamidi Sahner (2017) recently proposed a new method as a more robust approach. However, these methods are not designed to analyze the information sufficiency for a certain specific shock[20]. My interest has information sufficiency for one specific shock (i.e. monetary policy shock), due to which I adopt the approach of Forni et al. (2016) in the exercise below.

2.4.1 Informational deficiency

Definition Let $x_t = [x_{1,t}, ..., x_{n,t}]'$ denote a vector of n variables used in VAR, and assume that this vector has a moving average representation at DGP level:

$$x_t = \sum_{l=0}^{\infty} A_x(l) u_{t-l}$$

where $u_t = [u_{1,t}, ..., u_{q,t}]'$ are structural shocks ($q \times 1$ vector), and $A_x(l)$ represents an $n \times q$ matrix of impulse response functions. In principle, VAR with lag L decomposes the vector x_t to two orthogonal components:

$$x_t = P(x_t | H^x_{t-1}(L)) + \epsilon^{(L)}_t$$

[20]Their methods are designed to check the *fundamentalness* of the VAR system (or a similar concept to it). For detail, see Hansen and Sargent (1991), Lippi and Reichlin (1993), Lippi and Reichlin (1994), Chari et al. (2008), and Forni and Gambetti (2014).

where $H^x_{t-1}(L)$ is a closed linear space defined by x_{t-k} with $k = 1, ..., L$; and $P(a|b)$ represents a projection function of a onto b. Let $u_{i,t}$ (i-th component of u_t) be a shock of interest, and it can be written as

$$u_{i,t} \;=\; P(u_{i,t}|\epsilon_t^{(L)}) + e_{i,t}^{(L)}. \tag{2.14}$$

In (2.14), $e_{i,t}^{(L)}$ represents a discrepancy between the shock identified by VAR and the true shock, and this discrepancy is generated due to a deficiency of VAR's information set. On this basis, Forni et al. (2016) define informational deficiency of VAR for the shock $u_{i,t}$ by a fraction of the unexplained variance in the projection function (2.14):

$$\delta_i(L) = \frac{\sigma^2_{e_{i,(L)}}}{\sigma^2_{u_i}}.$$

Note that this measure is within a range between zero and one ($0 \leq \delta_i \leq 1$). We can regard VAR's information as being sufficient when δ_i is smaller than a pre-specified threshold M. There is no exact definition on which kind of value the threshold M should take, but Forni et al. (2016) propose setting $M = 0.05$ or 0.1 as the relevant level.

Formula Forni et al. (2016) derive a simple formula used to compute $\delta_i(L)$ as follows. First, Proposition $2'$ in their paper gives

$$P(u_{i,t}|\epsilon_t^{(L)}) = P(u_{i,t}|H^x_t(L)).$$

From this,

$$\begin{aligned} \delta_i(L) \;=\; \frac{\sigma^2_{e_{i,(L)}}}{\sigma^2_{u_i}} \;&=\; \frac{Var[u_{i,t} - P(u_{i,t}|e_t^{(L)})]}{\sigma^2_{u_i}} \\ &=\; \frac{Var[u_{i,t} - P(u_{i,t}|H^x_t(L))]}{\sigma^2_{u_i}} \end{aligned} \tag{2.15}$$

Using the definition $y_t \equiv [x'_t, ..., x'_{t-L}]'$, it is derived that

$$\begin{aligned} Var[u_{i,t} - P(u_{i,t}|H^x_t(L))] \;&=\; Var[u_{i,t} - u_{i,t}y'_t(y_t y'_t)^{-1}y_t] \\ &=\; \sigma^2_{u_i} - A^{(i)}_y(0)'\Sigma_y^{-1}A^{(i)}_y(0) \end{aligned} \tag{2.16}$$

where $A^{(i)}_y(0)$ represents the contemporaneous responses of y_t to $u_{i,t}$, and Σ_y is the variance covariance matrix of y_t. From (2.15) and (2.16), a formula to compute δ_i is given by

$$\delta_i(L) = 1 - A_y^{(i)}(0)'\Sigma_y^{-1}A_y^{(i)}(0)/\sigma_{u_i}^2. \tag{2.17}$$

Note that Σ_y is obtained by

$$\begin{cases} \Sigma_y &= \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_L \\ \Gamma_{-1} & \Gamma_0 & \cdots & \Gamma_{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{-L} & \cdots & \cdots & \Gamma_0 \end{pmatrix} \\ \\ \Gamma_k &= E[x_t x_{t-k}'] = \sum_{m=1}^{q} \sigma_{u_m}^2 \sum_{l=0}^{\infty} A_x^{(m)}(l)\left(A_x^{(m)}(l+k)\right)' \end{cases}$$

where q is the number of structural shocks, and $A_x^{(m)}(l)$ is the m-th column of the impulse response matrix $A_x(l)$.

Computation I compute the informational deficiency for monetary policy shock in the open-economy DSGE. For the VAR vector x_t, the following two designs are considered:

(i) $x_t = [R_t, Y_t, \pi_t]'$

(ii) $x_t = [R_t, (PC)_{1,t}, ..., (PC)_{K,t}]'$

where $(PC)_{j,t}$ is the j-th principal component obtained from all of the endogenous variables used in TVP/SV-FAVAR estimation (see Eq. (2.10)). The case (i) evaluates the deficiency of 3-variable VAR, whereas the case (ii) pertains to the deficiency of FAVAR. Note also that the lag of the VAR system is set equal to two ($L = 2$) in both cases. By Eq. (2.17), the informational deficiency is evaluated. The obtained results are shown in Table 2.2.

2.4.2 Discussion

As mentioned in 2.4.1, we can regard VAR/FAVAR's information as being sufficient if δ_i is smaller than a pre-specified threshold M. Using Forni et al. (2016)'s suggestion, the set is $M = 0.10$ in the discussion below.

Table 2.2 suggests that the 3-variable VAR ($x_t = [R_t, Y_t, \pi_t]'$) is *informationally deficient* in identifying monetary policy shock. In the case

Table 2.2: Informational deficiency of VARs for estimating a monetary policy shock in the open-economy DSGE. x_t represents the VAR vector. The lag of the VAR system is set equal to two ($L = 2$).

(i) $x_t = [R_t, Y_t, \pi_t]'$

Regime 1	Regime 2
0.459	0.510

(ii) $x_t = [R_t, (PC)_{1,t}, ..., (PC)_{K,t}]'$

	Regime 1	Regime 2
$K = 1$	0.639	0.489
$K = 2$	0.472	0.385
$K = 3$	0.446	0.283
$K = 4$	0.440	0.066
$K = 5$	0.148	0.064
$K = 6$	0.050	0.062
$K = 7$	0.047	0.005

of FAVAR ($x_t = [R_t, (PC)_{1,t}, ..., (PC)_{K,t}]'$), the information deficiency decreases as the number of principal components increases. Another finding from Table 2.2 is that the degree of deficiency can change across the different regimes. This change is prominent, especially when the information set of empirical models is not enough (see Table 2.2(ii) with small K). It should be emphasized that both informational deficiency and its variations diminish the ability of the empirical models to capture time variations in the monetary policy transmission. VAR easily suffers from both of these problems, but FAVAR can overcome them by virtue of the latent factors. This gives an essential explanation of why TVP/SV-FAVAR displayed good performance in the MC experiments and why TVP/SV-VAR did not.

Regarding the number of latent factors to achieve informational sufficiency, the implication from Table 2.2 is not necessarily consistent with the one in the MC experiments in Section 2.3. In the MC experiments, TVP/SV-FAVAR with $K = 2$ exhibited good performance, whereas Table 2.2 suggests that K should be six or more. To understand this inconsistency, it would be useful to recall that the exercise in 2.4.1 used principal components when examining the deficiency of FAVAR. As pointed out by Forni and Gambetti (2014), an exercise using principal components may overestimate the number of latent factors needed to achieve informational sufficiency, because the principal components may be imperfect estimates of the latent factors. Regarding this point, Stock and Watson (2010) argue that principal component estimation is based on cross-sectional averaging[21], whereas the Kalman-filter-based algorithm estimates the latent factors using information spanning the full sample period. In the MC experiments in Section 2.3, the latent factors in TVP/SV-FAVAR are estimated using the algorithm of Carter and Kohn (1994), and this algorithm computes the factors through Kalman filter and an additional backward recursive algorithm. Therefore, to the extent of relying on full period information, the obtained factors through this algorithm are expected to be better estimates than the principal components. From this, it can be understood why TVP/SV-FAVAR with a limited number of factors ($K \leq 3$) could display a good level of performance in the MC experiments.

2.5 An empirical application

In order to reinforce the findings in the MC experiments (see Section 2.3), I also conduct an empirical exercise with an application to Japan's data from the 1970s to the 2000s. The exercise exhibits different performances of TVP/SV-FAVAR and TVP/SV-VAR. Note that the analysis with TVP/SV-FAVAR was already conducted in Chapter 1. In what follows, the analysis with TVP/SV-VAR is newly added.

2.5.1 Estimation methodology

In TVP/SV-FAVAR, the common factors Z_t are defined by

$$Z_t \;=\; \{F_t, R_t\}' \;=\; \{F_t^1, ..., F_t^K, R_t\}' \tag{2.18}$$

where F_t and R_t represent the latent factors and monetary policy instrument, respectively. The number of latent factors is set equal to three ($K = 3$)[22]. In TVP/SV-VAR, the VAR vector is defined as follows:

$$X_t = \{(IIP)_t, (CPI)_t, R_t\}' \tag{2.19}$$

[21] Principal components at time t ($= (PC)_t$) are estimated by using only the information at the contemporaneous period, instead of relying on the information spanning the full sample period (i.e. $1 \leq t \leq T$).

[22] It is optimized by using the information criteria of Bai and Ng (2002). For detail, see 1.3.4.

where $(IIP)_t$ and $(CPI)_t$ represent the aggregate IIP and CPI, respectively. The estimation procedure for the two models is basically the same as in the MC exercise in Section 2.3, but the exceptions are that monetary policy shock is identified by the recursive scheme, and that the iteration times of the sampling algorithm is 30,000, with the first 27,000 draws removed as a burn-in ($(M, M_0) = (30,000, 27,000)$).

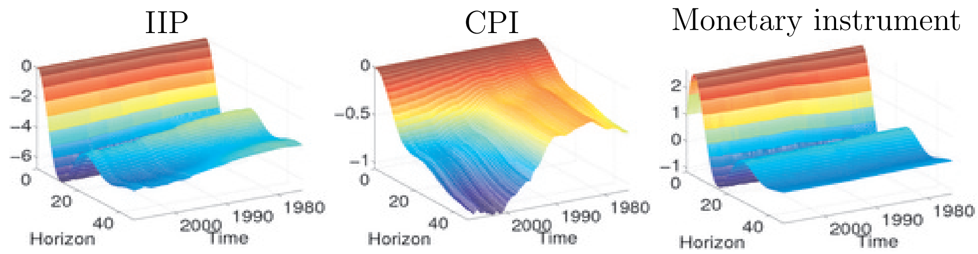
2.5.2 Results

TVP/SV-FAVAR and TVP/SV-VAR are estimated with the data set which I used in Chapter 1[23]. Fig. 2.7 shows the obtained impulse responses of the three major variables (IIP, CPI and monetary policy instrument) to monetary policy shock. Fig. 2.8 displays the responses of IIP and CPI at the long impulse horizon (h=48 months). In the case of TVP/SV-FAVAR, the response of CPI significantly strengthens during the early 1990s, and the timing of this change is associated with the bubble collapse, which happened in 1990-91 in Japan. On the other hand, TVP/SV-VAR indicates no evidence of time variation in the monetary policy effects. These results seem to reproduce the empirical literature on the Great Moderation, where FAVAR results suggest a significant change in monetary policy transmission but VAR does not identify the same.

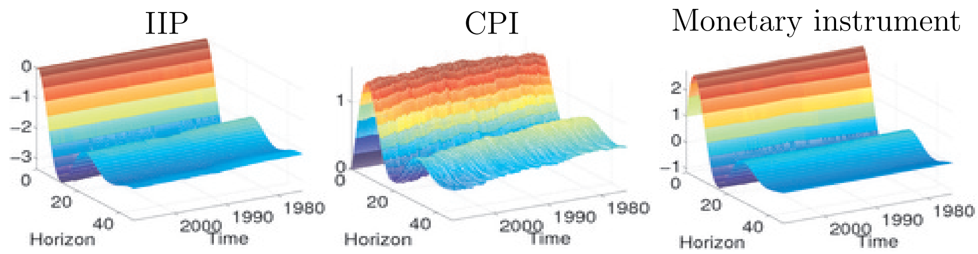
2.6 Conclusion

This study investigates the ability of TVP/SV-FAVAR to capture time variations in the transmission of monetary policy shock, in comparison to that of TVP/SV-VAR. The analysis is conducted mainly through MC-based experiments using the open-economy DSGE as the data-generating process. The experiments show that TVP/SV-VAR does not adequately detect time variation in the monetary policy transmission, but TVP/SV-FAVAR does. Subsequently, the experimental results are interpreted in terms of the information deficiency of the two empirical models. Using the technique of Forni et al. (2016), it is quantitatively confirmed that VAR does not contain sufficient enough information to estimate monetary policy shock. As for another finding, when the information set of the empirical model is far from being enough, the extent of the informa-

[23]For detail of the data set, see 1.3.3.



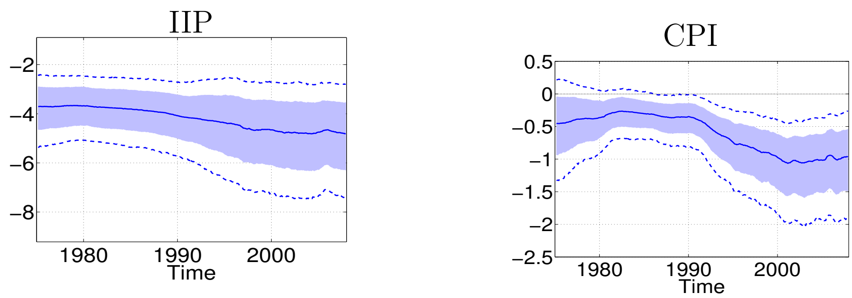
(a) TVP/SV-FAVAR



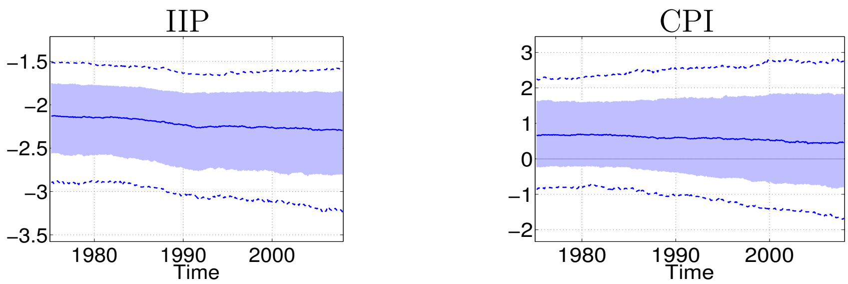
(b) TVP/SV-VAR

Figure 2.7: Time-varying median impulse responses to monetary policy shock from 1975 to 2007 in Japan. The estimated responses from (a) TVP/SV-FAVAR and (b) TVP/SV-VAR are depicted.

tional deficiency can vary significantly in accordance with a change in the underlying dynamics. It is generally expected that both informational deficiency and its variations diminish the ability of the empirical models to capture time variations in structural shock transmission. VAR easily suffers from both problems, but FAVAR can overcome them by virtue of using common latent factors. These findings provide an essential explanation as to why TVP/SV-FAVAR is more suitable than TVP/SV-VAR for identifying time variation in monetary policy effects. In order to reinforce the findings in the MC experiments, I also conduct the empirical exercise with an application to Japan's data from the 1970s to the 2000s. The obtained results seem to reproduce the findings in the literature of the Great Moderation, in the sense that FAVAR result suggests a significant variation in the monetary policy effects but VAR does not necessarily identify it.



(a) TVP/SV-FAVAR



(b) TVP/SV-VAR

Figure 2.8: Time-varying impulse responses at the impulse horizon $h = 48$ months from 1975 to 2007 in Japan. The estimated responses from (a) TVP/SV-FAVAR and (b) TVP/SV-VAR are depicted. The blue solid line, shaded area, and dotted lines represent the median, 16/84-percentiles, and 5/95-th percentiles of the estimated responses, respectively.

2.7 Appendices

2.7.1 Estimation of IRFs in MC experiments

In Section 2.3, the performance of the empirical models (TVP/SV-FAVAR and TVP/SV-VAR) is evaluated by using a cumulative impulse response to monetary policy shock at the long-run impulse horizon ($h = 24$). Cumulative impulse response at the long horizon indicates a total impact of monetary policy shock on each observed variable. In examining the time variation in the monetary policy transmissions, this measure would be a suitable one to check. However, it should be noticed that the estimated impulse response of the observed variables does not converge to zero at the long horizon (but to some non-zero level)[24]. In this case, the *cumulative* impulse response of the variables would not become flat at the long horizon, and therefore, the one at the long horizon would not represent the total effect of monetary policy shock. To avoid this issue, I apply a rough correction to the estimated impulse response of the variables. The correction scheme for each variable X_i is formulated as follows:

$$(Corr\ IRF)_{t,h}^{X_i} = (IRF)_{t,h}^{X_i} + C_i \tag{2.20}$$

where $(IRF)_{t,h}^{X_i}$ is an estimated impulse response of the variable X_i to monetary policy shock at time t for horizon h. The correction factor C_i is set equal to some of the values, such that the impulse response at Regime 1 ($t \leq 125$) converges to zero at the long horizon region ($h = 24$-48) on average. Note that C_i is constant for each observable X_i, and that it does not depend on either time t or horizon h. Therefore, it should be emphasized that this correction has no impact on the time variation in the estimated cumulative IRF of the variable X_i.

[24]It is likely to occur, especially when *log-difference transformation* is applied to the variables. If this transformation is adopted, the *cumulative* impulse response of the log-differenced variable corresponds to the impulse response of the observed variable (*non-transformed* variable). Due to the stationarity of the system, the impulse response (non-cumulative response) of the log-differenced variable tends to converge to zero at the long horizon, but this is not necessarily the case for the cumulative response. This means that the estimated impulse response of the observed variable (non-transformed variable) does not necessarily converge to zero at the long horizon.

2.7.2 Optimization of K

To optimize the number of latent factors K in TVP/SV-FAVAR, I conduct an MC exercise in the three cases of K ($K \leq 3$). As for the estimation procedure of the models, see Section 2.3.

Fig. 2.9 shows the estimated median cumulative IRFs to monetary policy shock across the 100 experiments. When $K \geq 2$ (Figs. 2.9(b) and (c)), the long-run responses of output and inflation strengthen around $t = 125$. These results are consistent in terms of both the timing and direction of the variation, with the true change in the underlying DSGE. However, in the case of $K = 1$ (Fig. 2.9(a)), the responses of both output and inflation display a strange result. The response of the output does not indicate a clear variation across the two regimes, and that of the inflation decreases around $t = 125$.

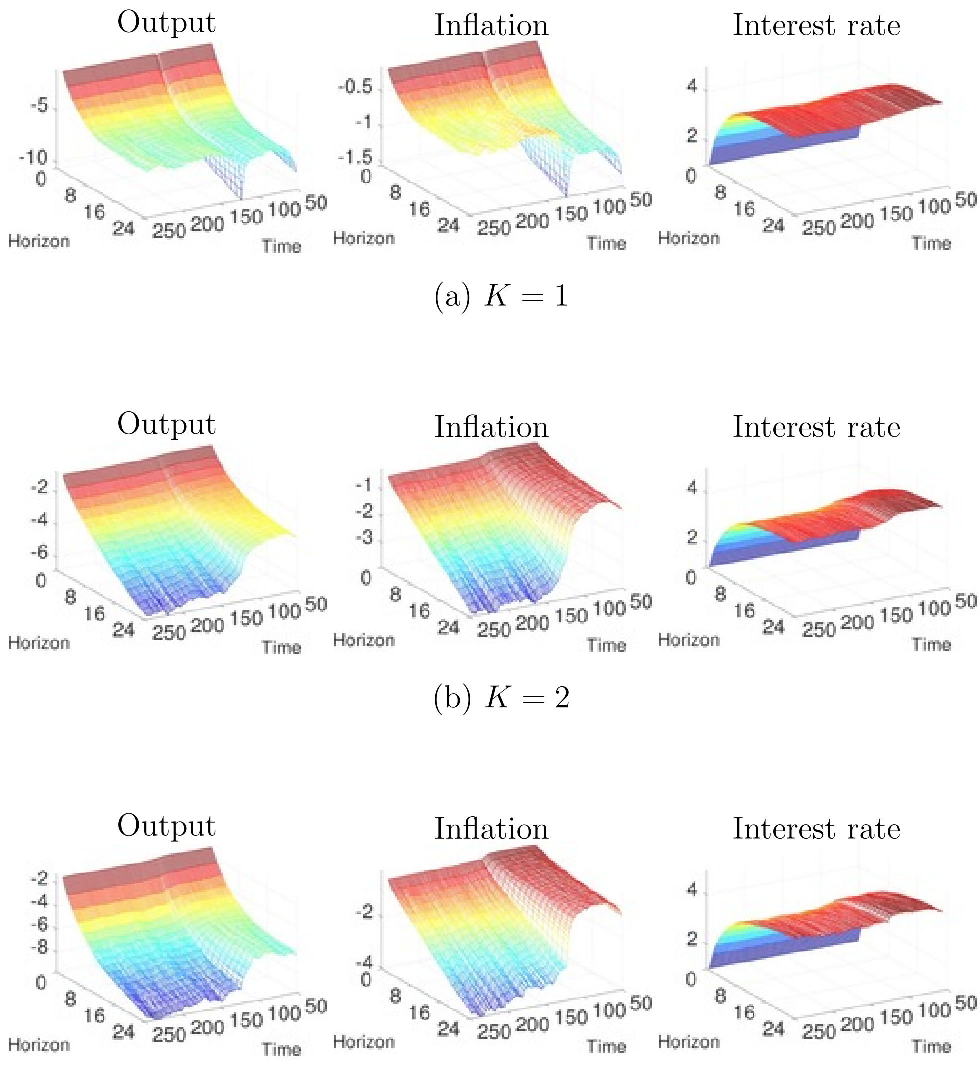
As for the responses of output and inflation, I also check their time variation at a fixed horizon of $h = 24$, as shown in Fig. 2.10. In all of the plots, the estimated response is compared with the true one as indicated by the red line. When $K = 1$, the estimated response clearly deviates from the true one (in both terms of output and inflation), and the gap between them cannot be explained by estimation uncertainty. On the contrary, FAVARs with $K \geq 2$ exhibit that the estimated result is consistent with the truth within the bounds of estimation uncertainty. In another finding, the uncertainty bands in the case of $K = 3$ tend to be larger than the ones with $K = 2$. This reflects the fact that uncertainty in model estimation is enlarged as the number of latent factors increases.

Based on the above observations, I choose $K = 2$ as the case yielding TVP/SV-FAVAR's best performance.

2.7.3 Exercise with Smets-Wouters model (2007)

As a further robustness check, I conduct an MC exercise using Smets-Wouters model (2007) as the DGP.

Smets-Wouters model (2007) As well as the open-economy DSGE, the Smets-Wouters model (2007) contains many shocks and frictions that affect households and firms. Its key features include nominal price and wage settings, habit information in consumption, and investment adjust-

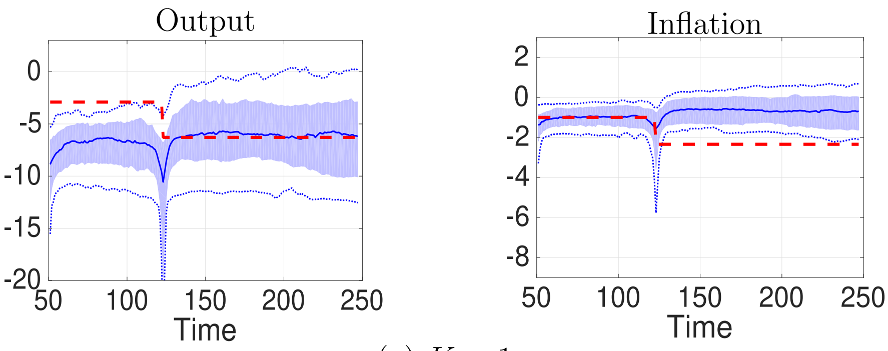


(a) $K = 1$

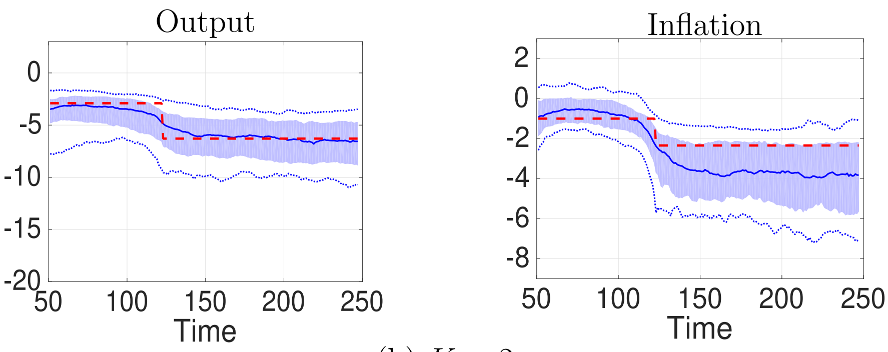
(b) $K = 2$

(c) $K = 3$

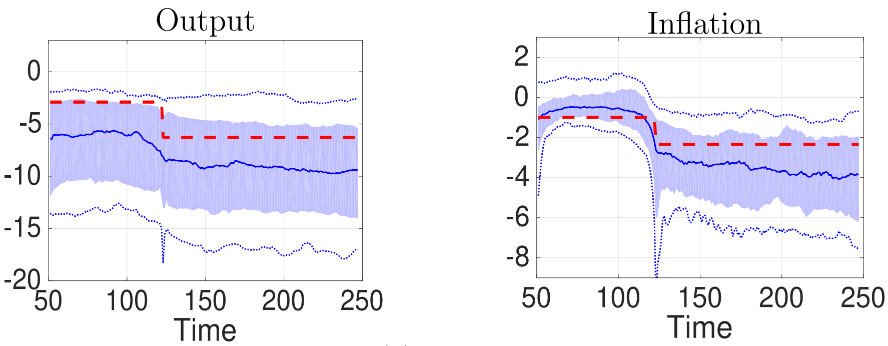
Figure 2.9: Median cumulative impulse responses to monetary policy shock across the 100 pseudo-experiments using TVP/SV-FAVARs with (a) $K = 1$, (b) $K = 2$ and (c) $K = 3$.



(a) $K = 1$



(b) $K = 2$



(c) $K = 3$

Figure 2.10: Cumulative impulse responses at the impulse horizon $h = 24$. The responses are obtained by the 100 pseudo-experiments using TVP/SV-FAVARs with (a) $K = 1$, (b) $K = 2$ and (c) $K = 3$. The blue solid line, shaded area, and dotted lines represent the median, 16/84-percentiles, and 5/95-th percentiles of the estimated responses, respectively. The red dashed line indicates the true response in the DSGE.

Table 2.3: Endogenous variables in the Smets-Wouters model (2007).

Notation	Definition
Y_t	output
C_t	consumption
L_t	labor (hours worked)
R_t	nominal interest rate
π_t	inflation
W_t	real wage
$\widetilde{\mu}_t^w$	wage mark-up
$\widetilde{\mu}_t^p$	price mark-up
I_t	investment
$\widetilde{q}_t$	value of capital stock
K_t	capital installed
K_t^s	capital used in production
R_t^k	rental rate of capital
$\widetilde{z}_t$	capital utilization costs

ment costs. On the basis of the works by Smets and Wouters (2003) and Christiano et al. (2005), Smets and Wouters (2007) extend the monetary business cycle model so then it is consistent with a balanced steady-state growth path driven by deterministic technological progress. The model consists of seven exogenous shocks and fourteen endogenous variables. All of the endogenous variables are listed in Table 2.3.

In simulating the model, the sample size is set equal to 250. As for calibration, the estimates of Smets and Wouters (2007) are used. For each of Regimes 1 ($1 \leq t \leq 125$) and 2 ($126 \leq t \leq 250$), their estimates with 1966-1979 and 1984-2004 subsamples are used, respectively. The simulation procedure is the same as in the exercise with the open-economy DSGE (baseline exercise) as described in Section 2.3. First, the model is solved separately by the regimes. Then, by simulating the models 100 times, 100 pseudo-data samples are prepared. In each simulation, 500 observations are generated for each regime, and the first 375 observations are discarded. Fig. 2.11 displays the theoretical IRFs of output, inflation, and interest rate to monetary policy shock. As for the key features, the response of output varies remarkably across the two regimes, whereas those of inflation and interest rate do not vary significantly.

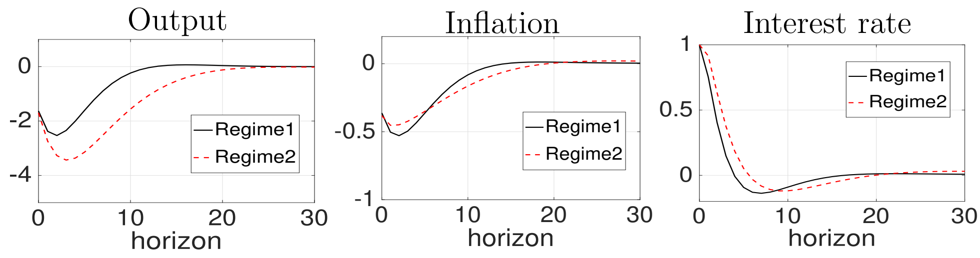


Figure 2.11: Theoretical impulse response functions (IRFs) of output, inflation, and interest rate to monetary policy shock in the Smets-Wouters model (2007). The black solid line and the red dashed line represent the responses in the cases of Regimes 1 and 2, respectively.

Pseudo-experiments and results The observed vectors in TVP/SV-FAVAR and TVP/SV-VAR are defined by:

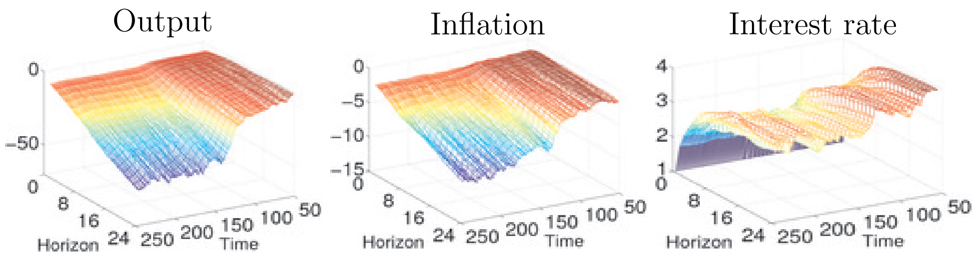
$$\begin{aligned} X_t &= \{\tilde{z}_t, C_t, K_t^s, W_t, Y_t, \pi_t, L_t, R_t^k\}' \quad &\text{(for TVP/SV-FAVAR)} \\ X_t &= \{R_t, Y_t, \pi_t\}' \quad &\text{(for TVP/SV-VAR)} \end{aligned} \tag{2.21}$$

For the definition of the variables, see Table 2.3. The identification scheme and estimation procedure are the same as in the baseline exercise, but the exception is that the number of latent factors in TVP/SV-FAVAR is set equal to three ($K = 3$).

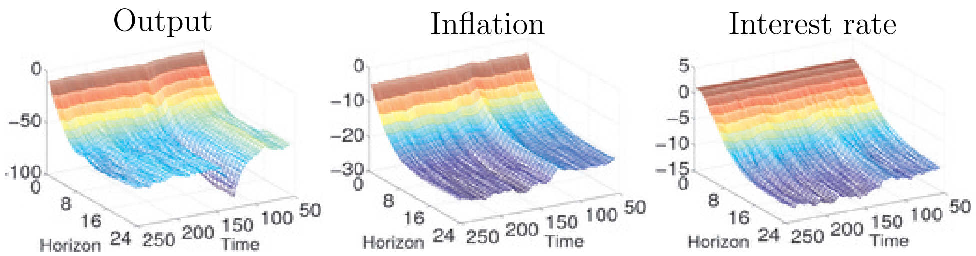
Figs. 2.12 and 2.13 display the obtained impulse responses to monetary policy shock. In the case of TVP/SV-FAVAR, the estimated response of output strengthens around $t = 125$, while those of inflation and interest rate do not clearly vary across the whole of the study period[25]. In the case of TVP/SV-VAR, a clear time variation is not observed in any of the three variables. Moreover, it is especially strange that the response of the interest rate is far below zero at a long horizon, which also indicates that TVP/SV-VAR fails to identify monetary policy shock correctly.

Interpretation The above exercise has provided the same finding as in the baseline exercise (with the open-economy DSGE), indicating that TVP/SV-VAR does not adequately capture time variations in monetary policy transmission, but TVP/SV-FAVAR does. To interpret this result,

[25]In Fig. 2.12(a), it looks like the response of inflation strengthens around $t = 125$, but Fig. 2.13(a) shows that this change is not statistically significant.

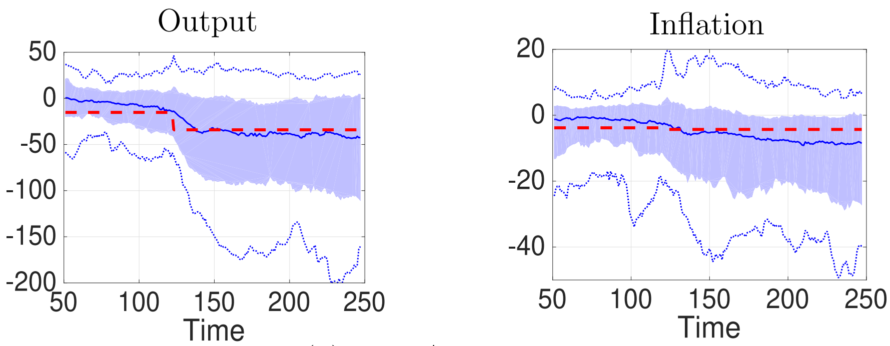


(a) TVP/SV-FAVAR

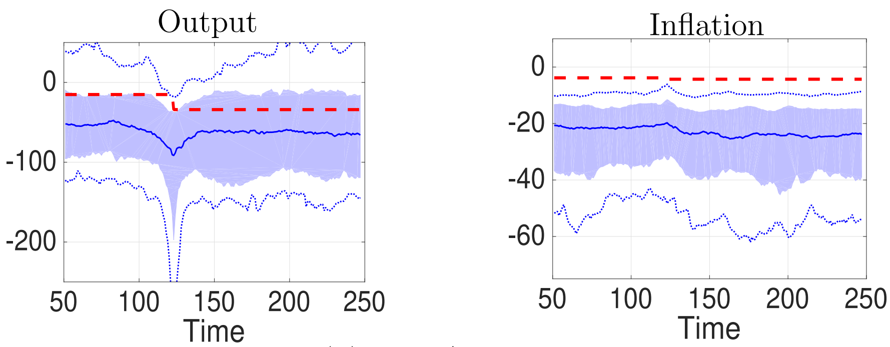


(b) TVP/SV-VAR

Figure 2.12: Time-varying median cumulative impulse responses to monetary policy shock across the 100 pseudo-experiments using the Smets-Wouters model (2007). The estimated responses from (a) TVP/SV-FAVAR and (b) TVP/SV-VAR are depicted.



(a) TVP/SV-FAVAR



(b) TVP/SV-VAR

Figure 2.13: Time-varying cumulative impulse responses at the impulse horizon $h = 24$ with the Smets-Wouters model (2007). The responses are obtained by the 100 pseudo-experiments using (a) TVP/SV-FAVAR and (b) TVP/SV-VAR. The blue solid line, shaded area, and dotted lines represent the median, 16/84-percentiles, and 5/95-th percentiles of the estimated responses, respectively. The red dashed line indicates the true response in the DSGE.

Table 2.4: Informational deficiency of VARs for estimating a monetary policy shock in the Smets-Wouters model (2007). x_t represents the VAR vector. The lag of the VARs is set equal to two ($L = 2$).

(i) $x_t = [R_t, Y_t, \pi_t]'$

Regime 1	Regime 2
0.345	0.229

(ii) $x_t = [R_t, (PC)_{1,t}, ..., (PC)_{K,t}]'$

	Regime 1	Regime 2
$K = 1$	0.581	0.406
$K = 2$	0.521	0.323
$K = 3$	0.473	0.305
$K = 4$	0.393	0.272
$K = 5$	0.188	0.176
$K = 6$	0.004	0.003
$K = 7$	0.002	0.002

informational deficiency[26] is calculated for the following two cases of VAR vector x_t:

(i) $x_t = [R_t, Y_t, \pi_t]'$

(ii) $x_t = [R_t, (PC)_{1,t}, ..., (PC)_{K,t}]'$

where $(PC)_{j,t}$ is the j-th principal component obtained from all of the endogenous variables used in the TVP/SV-FAVAR estimation (see Eq. (2.21)).

The results are shown in Table 2.4. In the case of 3-variable VAR, the information set is far from begin enough, whereas in the case of FAVAR, information sufficiency is achieved by virtue of using the latent factors. Based on the same arguments as in Section 2.4, it can be understood why TVP/SV-FAVAR displayed a good performance in the previous subsection and why this is not the case with TVP/SV-VAR.

[26]For the definition of informational deficiency, see Section 2.4.

References for Chapter 2

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70:191–221.

Baumeister, C. and Benati, L. (2013). Unconventional monetary policy and the Great Recession: Estimating the macroeconomic effects of a spread compression at the zero lower bound. *International Journal of Central Banking*, 9:165–212.

Baumeister, C., Liu, P., and Mumtaz, H. (2013). Changes in the effects of monetary policy on disaggregate price dynamics. *Journal of Economic Dynamics and Control*, 37:543–560.

Baumeister, C. and Peersman, G. (2013). Time-varying effects of oil supply shocks on the US economy. *American Economic Journal: Macroeconomics*, 5:1–28.

Benati, L. and Surico, P. (2009). VAR analysis and the Great Moderation. *American Economic Review*, 99:1636–1652.

Bernanke, B., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A Factor-augmented Vector Autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120:387–422.

Bianchi, F., Mumtaz, H., and Surico, P. (2009). The Great Moderation of the term structure of interest rates. *Journal of Monetary Economics*, 56:856–871.

Blanchard, O. and Simon, J. (2001). The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity*, 32:135–174.

Canova, F. (2006). You can use VARs for structural analyses. A comment to VARs and the Great Moderation. mimeo, UPF.

Canova, F. and Gambetti, L. (2010). Do expectations matter? The Great Moderation revisited. *American Economic Journal: Macroeconomics*, 2:183–205.

Canova, F. and Hamidi Sahner, M. (2017). Are small-scale SVARs useful for business cycle analysis? Revisiting non-fundamentalness. *Journal of the European Economic Association*, 16:1069–1093.

Canova, F. and Nicol, G. D. (2002). Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics*, 49:1131–1159.

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–553.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2008). Are structural VARs with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics*, 55:1337–1352.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113:1–45.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics*, 8:262–302.

Ellis, C., Mumtaz, H., and Zabczyk, P. (2014). What lies beneath? A time-varying FAVAR model for the UK transmission mechanism. *Economic Journal*, 0:668–699.

Fernandez-Villaverde, J. and Rubio-Ramirez, J. F. (2008). How structural are structural parameters? In *NBER Macroeconomics Annual 2007*, volume 22, pages 83–137. National Bureau of Economic Research, Inc.

Forni, M. and Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics*, 66:124–136.

Forni, M., Gambetti, L., and Sala, L. (2016). VAR information and the empirical validation of DSGE models. CEPR Discussion Papers 11178.

Gali, J. and Gambetti, L. (2009). On the sources of the Great Moderation. *American Economic Journal: Macroeconomics*, 1:26–57.

Gambetti, L., Pappa, E., and Canova, F. (2008). The structural dynamics of US output and inflation: What explains the changes? *Journal of Money, Credit, and Banking*, 40:369–388.

Giannone, D. and Reichlin, L. (2006). Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4:455–465.

Graeve, F. D. (2017). A structural evaluation of VARs with time-varying parameters and stochastic volatility. Working Paper.

Hansen, L. and Sargent, T. J. (1991). Two problems in interpreting vector autoregressions. In Hansen, L. and Sargent, T., editors, *Rational Expectations Econometrics*, pages 77–119. Boulder : Westview.

Jacquier, E., Polson, N., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–417.

Justiniano, A. and Preston, B. (2010). Monetary policy and uncertainty in an empirical small open-economy model. *Journal of Applied Econometrics*, 25:93–128.

Kapetanios, G., Mumtaz, H., Stevens, I., and Theodoridis, K. (2012). Assessing the economy-wide effects of quantitative easing. *Economic Journal*, 122:F316–F347.

Kilem, M., Kriwoluzky, A., and Sarferaz, S. (2015). On the low-frequency relationship between public deficits and inflation. *Journal of Applied Econometrics*, 31:566–583.

Lippi, M. and Reichlin, L. (1993). The dynamic effects of aggregate demand and supply disturbances : Comment. *American Economic Review*, 83:644–652.

Lippi, M. and Reichlin, L. (1994). VAR analysis, non-fundamental representation, Blaschke matrices. *Journal of Econometrics*, 63:307–325.

Monacelli, T. (2005). Monetary policy in a low pass-through environment. *Journal of Money, Credit, and Banking*, 37:1047–1066.

Perez-Quiros, G. and McConnell, M. M. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review*, 90:1464–1476.

Primiceri, G. E. (2005). Time varying structural vector autoregression and monetary policy. *Review of Economic Studies*, 72:821–852.

Sims, C. A. (2002). Solving linear rational expectations models. *Computational Economics*, 20:1–20.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy? *American Economic Review*, 96:54–81.

Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the Euro area. *Journal of the European Economic Association*, 1:1123–1175.

Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97:586–606.

Stock, J. H. and Watson, M. W. (2002). Has the business cycle changed and why? *In M. Gertler and K. Rogoff, eds. (2003), NBER Macroeconomics Annuals 2002*.

Stock, J. H. and Watson, M. W. (2010). Dynamic factor models. In Clements, M. and Hendry, D., editors, *Oxford Handbook of Economic Forecasting*. Oxford University Press.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52:381–419.

Yamamura, T. (2017). Changes in monetary policy effects through the bubble burst and the (near-)zero interest rate regime in Japan. Working Paper.

Chapter 3

Estimating Large Panels with Unknown Group Multifactor Structures

3.1 Introduction

Recently there has been an increasing literature on panel data models with multiple unobserved common factors. By adopting time-varying common factors that affect the individuals with individual-specific factor loadings, the models incorporate both individual and time effects. As a notable feature, these individual and time effects are interacted multiplicatively, which introduces so-called interactive fixed effects (IFEs). With such a multiplicative form, the models can capture unobserved heterogeneity more flexibly than the ones with additive individual and time effects. Due to this flexibility, individual fixed effects have become a powerful tool in the econometrics literature.

In the literature using IFEs, many works consider the models with additional individual-specific regressors, and focus on a consistent estimation of the corresponding regression coefficients. Such studies fall into two categories depending on how to treat the unobserved factor structure. In the first category, Pesaran (2006) proposes the common correlated effects (CCE) estimator. In his approach, he uses cross sectional averages of the dependent variable and the individual-specific regressors, as a proxy for the common factors. The second category relies on the principal components analysis (PCA) approach. As a pioneering work, Bai (2009) derives the asymptotic properties of the least squares estimator, and Moon and Weidner (2017) extend them to the models with predetermined regressors. Moon and Weidner (2015) consider the case of unknown number of factors. Lu and Su (2016) propose an adaptive Lasso method for selecting the relevant regressors and determining the number of factors. Ando and Bai (2016) propose a model with grouped factor structure, in which the group membership and the number of groups are left unspecified.

In this chapter, the Pesaran (2006)'s approach is extended to a model with grouped factor structure (grouped IFEs). To understand an intuition of such a discrete structure, it may be useful to consider multi-country data. In multi-country data, there could be country (individual-specific) and region (group-specific) variations. In this case, the grouped factor structure captures the region-level variation.

Throughout this chapter's study, the parameters of interest are the (homogeneous) coefficients for the individual-specific regressors. The novelty of this study is that a Lasso technique is adopted in the regression loss function. To identify a grouped structure of IFEs, the convex clustering method is applied to the factor loadings. Using this approach, we aim at proposing a more precise estimator than the one which does not deal with group identification.

The group identification in our approach includes an estimation of unknown group memberships. Regarding this challenging problem, pioneering works are provided by, for example, Sun (2005), Lin and Ng (2012), and Bohomme and Manresa (2015), but none of their works consider factor error structures. Ando and Bai (2016) introduce grouped factor structure with unknown membership, but they do not rely on a Lasso approach for the group identification.

The convex clustering approach has been recently proposed by Hocking et al. (2011). To identify the grouped fixed effects, they introduce an ℓ_1-constraint on the pair-wise difference of the individual fixed effects (i.e. Lasso approach). Further analysis and modifications are also provided by Zhu et al. (2014), Tan and Witten (2015), Radchenko and Mukherjee (2017), and Gu and Volgushev (2018). It should be noted that all of those works are based on the fixed effects model. Our study focuses on a model with grouped factor structure (i.e. grouped *interactive* fixed effects), which is an extended version of the grouped fixed effects model.

It should be also mentioned that our approach relies on a modified Lasso, which is called *Square-root Lasso*. The Square-root Lasso approach is originally proposed by Belloni et al. (2011). As an attractive feature, it handles unknown scale of the noise σ_ε (i.e. the penalty level can be chosen without the knowledge of σ_ε)[1].

In this chapter, the performance of the square-root Lasso estimator is examined in comparison with two other estimators. One is the existing estimator proposed by Pesaran (2006). This is an estimator based on a pooling approach, and it does not deal with identification of grouped factor structure. The second one is what we call an *oracle estimator*, which is an infeasible estimator based on the true grouping of the factor structure.

The analysis procedure can be summarized as follows. First, we focus on a series of theoretical discussion. It is revealed what sort of condition at the population level makes the oracle estimator more efficient than the existing estimator. Using a simplified model framework, we also discuss that the Lasso estimator is likely to acquire *oracle property* (i.e. the estimator behaves as if it knows the group structure). To examine small sample properties of the Lasso estimator, we subsequently conduct Monte-Carlo (MC) based experiments. In a wide range of model parameter's space, the Lasso estimator improves RMSE (root mean squared error) of the existing estimator to a substantial extent, and it also exhibits a comparable performance to the oracle estimator. Through the experimental results, we also discuss the conditions under which the performance of the estimator is deteriorated in comparison with the oracle estimator.

This chapter is organized as follows. In Section 3.2, the model and estimation methodology are described. Section 3.3 discusses theoretical properties of the estimators. In Section 3.4, the Monte Carlo experiments are conducted. After discussing a possibility of model extension in Section 3.5, Section 3.6 provides a conclusion. As to the properties of the estimators, some theoretical proofs are also given in Section 3.7.

[1] As other features, it can also handle heteroskedasticity and non-Gaussianity of the noise, but these matters are beyond the scope of this study.

3.2 Model and estimation procedures

3.2.1 Model and notations

Consider the following data generating process (DGP):

$$\begin{aligned} y_{it} &= \beta' X_{it} + \gamma_i' F_t + \varepsilon_{it}, \\ X_{it} &= \Gamma_i' F_t + V_{it}. \end{aligned} \tag{3.1}$$

where y_{it} is the dependent variable and X_{it} is a $k \times 1$ covariate, with $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The parameter of interest is the $k \times 1$ common slope vector β. In this model, F_t is a $m \times 1$ vector of unobserved factors, and $[\varepsilon_{it}, V_{it}']'$ is a vector of independently and identically distributed (i.i.d) disturbance terms. The disturbances are assumed to be independent of $\{[X_{it}', F_t'], (i,t) \in \mathbb{N}^2\}$. It is also assumed that the γ_i's can take J_N values, where J_N is unknown. It follows that there exists a partition $\mathcal{G} = \{G_1, \ldots, G_{J_N}\}$ of $\{1, \ldots, N\}$ such that, for any element G of the partition $\gamma_i = \overline{\gamma}_G$ for all i in G. Let N_j be the number of indices i in G_j, $j = 1, \ldots, J_N$. Let us also suppose that N is very large but J_N is finite ($J_N \ll N$).

It should be also noticed that the factors (F_t) and the loadings (γ_i and Γ_i) are not necessarily identifiable[2]. To identify them, we impose the following restrictions.

$$\begin{cases} \displaystyle\sum_{t=1}^{T} F_t F_t' = T I_m \\ \left(\displaystyle\sum_{i=1}^{N} \gamma_i \gamma_i'\right) \text{ and } \left(\displaystyle\sum_{i=1}^{N} \Gamma_i^{(j)} \Gamma_i^{(j)\prime}\right) \text{ are diagonal } (1 \leq j \leq k) \,. \end{cases} \tag{3.2}$$

Note that $\Gamma_i^{(j)}$ $(1 \leq j \leq k)$ is a $m \times 1$ vector, which is defined in the $m \times k$ matrix Γ_i $\left(\text{i.e. } \Gamma_i = [\Gamma_i^{(1)} \ ... \ \Gamma_i^{(k)}]\right)$.

Our estimation procedure uses the common correlated effects (CCE) procedure considered in Pesaran (2006), who uses the averaged observa-

[2]For identification problem in factor models, see, for example, Bai (2009).

tions as a proxy for the factor. The two model equations imply

$$\begin{aligned}\underset{(k+1)\times 1}{Z_{it}} &= \begin{bmatrix} y_{it} \\ X_{it} \end{bmatrix} = \begin{bmatrix} \left(\beta'\Gamma_i' + \gamma_i'\right) F_t + \beta' V_{it} + \varepsilon_{it} \\ \Gamma_i' F_t + V_{it} \end{bmatrix} \\ &= \underset{(k+1)\times m}{C_i'} \cdot \underset{m\times 1}{F_t} + U_{it}\end{aligned} \tag{3.3}$$

where $U_{it} = [\beta' V_{it} + \varepsilon_{it}, V_{it}']'$ and

$$C_i = [\Gamma_i \beta + \gamma_i, \Gamma_i] = [\gamma_i, \Gamma_i] \cdot \begin{bmatrix} 1 & \mathbf{0}_{1\times k} \\ \beta & \mathrm{Id}_k \end{bmatrix}.$$

The CCE procedure is based on the individual average of the Z_{it}'s

$$\overline{Z}_t = \frac{1}{N}\sum_{i=1}^{N} Z_{it} \tag{3.4}$$

which is used as a proxy for the unobserved factor. Indeed, $\overline{Z}_t = \overline{C}' F_t + \overline{U}_t$, so that assuming that $\overline{CC}'$ has an inverse gives

$$F_t = \left(\overline{CC}'\right)^{-1} \overline{C}\left(\overline{Z}_t - \overline{U}_t\right).$$

Substituting in the model equation for y_{it} then gives

$$\begin{aligned} y_{it} &= \beta' X_{it} + \gamma_i' \left(\overline{CC}'\right)^{-1} \overline{CZ}_t + \varepsilon_{it} - \gamma_i' \left(\overline{CC}'\right)^{-1} \overline{CU}_t \\ &= \beta' X_{it} + \alpha_i' \overline{Z}_t + \eta_{it} \end{aligned} \tag{3.5}$$

where $\alpha_i = \overline{C}' \left(\overline{CC}'\right)^{-1} \gamma_i$ is a $(k+1)\times 1$ vector, and η_{it} is defined by

$$\begin{aligned} \eta_{it} &= \varepsilon_{it} - \gamma_i' \left(\overline{CC}'\right)^{-1} \overline{CU}_t \\ &= \varepsilon_{it} - \alpha_i' \overline{U}_t \end{aligned} \tag{3.6}$$

Throughout this study, Eq. (3.5) is used as a regression model. As an issue, this equation suffers from an endogenous problem due to a correlation between $\overline{Z}_t$ and η_{it}. In Section 3.4, we discuss the conditions under which this issue can be non-negligible or serious.

3.2.2 Existing estimator

In Pesaran (2006)'s methodology, β can be estimated by the following pooling approach:

$$\widehat{\beta}_{CCE} = \left(\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^{(p)}\widetilde{y}_{it}^{(p)\prime}\right) \tag{3.7}$$

where

$$\begin{aligned}\widetilde{X}_{it}^{(p)} &= X_{it} - \left(\sum_{s=1}^{T}X_{is}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t \\ \widetilde{y}_{it}^{(p)} &= y_{it} - \left(\sum_{s=1}^{T}y_{is}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t\end{aligned} \tag{3.8}$$

This is called a common correlated effects pooled (CCEP) estimator. Note that this estimator is not oriented to a group identification in the factor structure.

3.2.3 Oracle estimator

Before proposing the Lasso-based estimator, let us introduce an *oracle estimator*. This is an infeasible estimator based on the true grouping of the factor loadings. The estimator is defined in the following regression equation:

$$y_{it} = \beta' X_{it} + \alpha_{g(i)}'\overline{Z}_t + \eta_{it} \tag{3.9}$$

where $g(i) = \sum_{j=1}^{J_N} G_j \cdot \mathbb{I}(i \in G_j)$. The formula for this estimator is

$$\widehat{\beta}_{CCE}^{(OR)} = \left(\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\widetilde{y}_{it}\right) \tag{3.10}$$

with

$$\begin{cases}\widetilde{y}_{it} \equiv y_{it} - \dfrac{1}{N_{g(i)}}\cdot\left(\displaystyle\sum_{s=1}^{T}\sum_{k\in g(i)}y_{ks}\overline{Z}_s'\right)\left(\displaystyle\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t \\ \widetilde{X}_{it} \equiv X_{it} - \dfrac{1}{N_{g(i)}}\cdot\left(\displaystyle\sum_{s=1}^{T}\sum_{k\in g(i)}X_{ks}\overline{Z}_s'\right)\left(\displaystyle\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t\end{cases} \tag{3.11}$$

Throughout this study, let us call this estimator as *oracle CCE estimator*. For the derivation of Eq. (3.10), see 3.7.1.

3.2.4 Square-root classifier LASSO (SRC-LASSO)

The Lasso-based estimator is defined as follows. The least squares objective function is

$$\widehat{Q}^2(\boldsymbol{a}, b) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - b' X_{it} - a_i' \overline{Z}_t\right)^2$$

where $\boldsymbol{a} = [a_{1,} \ldots, a_n]'$. Minimizing $\widehat{Q}^2(\boldsymbol{a}, b)$ gives a CCE estimator that ignores the group structure. We therefore penalize for the variation of the a_i's. For $W = [w_{ik}, 1 \leq i, k \leq N]$ with $w_{ik} \geq 0$, set

$$\begin{aligned} \mathfrak{p}_W(\boldsymbol{a}) &= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} w_{i_1,i_2} \left|a_{i_1} - a_{i_2}\right| \\ &= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{l=1}^{r} w_{i_1,i_2}^{(l)} \left|a_{i_1}^{(l)} - a_{i_2}^{(l)}\right| \end{aligned} \tag{3.12}$$

where $r \equiv k + 1$. Note that $|\cdot|$ is a vector norm, and that a_{i_1} (and a_{i_2}) is an $r \times 1$ vector. The function $\mathfrak{p}_W(\cdot)$ is a LASSO type penalty which is 0 when the a_i's do not depend upon i and forces them to group. The square root classifier LASSO (or SRC-LASSO) estimator is

$$\left\{\widehat{\boldsymbol{\alpha}}\left(\widehat{W}, \widehat{\lambda}\right), \widehat{\beta}\left(\widehat{W}, \widehat{\lambda}\right)\right\} = \arg\min_{\boldsymbol{a}, b} \left\{\widehat{Q}(\boldsymbol{a}, b) + \widehat{\lambda} \cdot \mathfrak{p}_{\widehat{W}}(\boldsymbol{a})\right\}$$

for possibly random weight matrix $\widehat{W}$ and penalty parameter $\widehat{\lambda}$.

3.2.5 Estimation procedure

We use a two step choice of the weight $\widehat{W}$ of the SRC-LASSO estimator. In the first stage, the weights are defined by the CCEP approach:

$$\widehat{w}_{1i_1i_2}^{(l)} = \mathbb{I}(|\widehat{\alpha}_{i_1,CCE}^{(l)} - \widehat{\alpha}_{i_2,CCE}^{(l)}| < \delta)$$

where $\widehat{\alpha}_{i,CCE} = (\overline{Z}'\overline{Z})^{-1}\overline{Z}'\widehat{y}_{i,CCE}$ with $\widehat{y}_{i,CCE} = y_i - \widehat{\beta}_{CCE}' X_i$. Using this, the first SRC-LASSO estimator is

$$\begin{aligned} &\left\{\widetilde{\boldsymbol{\alpha}}_{lasso}, \widetilde{\beta}_{lasso}\right\} = \arg\min_{\boldsymbol{a}, b} \left\{\widehat{Q}(\boldsymbol{a}, b) + \lambda_1 \cdot P_1(\boldsymbol{a})\right\} \\ &P_1(\boldsymbol{a}) = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{l=1}^{r} \widehat{w}_{1i_1i_2}^{(l)} \left|a_{i_1}^{(l)} - a_{i_2}^{(l)}\right| \end{aligned}$$

where λ_1 is some deterministic sequence going to 0 at a suitable rate.

The role of the first-stage is to form groups, which are used in a second-stage to refine the weight penalty which are set to 0 when i and j do not belong to the same estimated group. The second-stage estimator is written as

$$\left\{\widehat{\boldsymbol{\alpha}}_{lasso}, \widehat{\beta}_{lasso}\right\} = \arg\min_{\boldsymbol{a},b} \left\{\widehat{Q}\left(\boldsymbol{a}, b\right) + \lambda_2 \cdot P_2\left(\boldsymbol{a}\right)\right\}.$$

where

$$\widehat{P}_2\left(\boldsymbol{a}\right) = \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{l=1}^{r} \widehat{\omega}_{2i_1i_2}^{(l)} \left| a_{i_1}^{(l)} - a_{i_2}^{(l)} \right|,$$

$$\widehat{\omega}_{2i_1i_2}^{(l)} = \frac{\mathbb{I}\left(\widetilde{\alpha}_{i_2}^{(l)} = \widetilde{\alpha}_{i_1}^{(l)}\right)}{\sqrt{\widetilde{\nu}_{i_1}^{(l)}} \cdot \sqrt{\widetilde{\nu}_{i_2}^{(l)}}}, \quad \widetilde{\nu}_i^{(l)} = \sum_{i'=1}^{N} \mathbb{I}\left(\widetilde{\alpha}_{i'}^{(l)} = \widetilde{\alpha}_i^{(l)}\right).$$

3.2.6 Assumptions

This subsection collects some assumptions on the DGP.

Assumption 1
(i) For any i and t, ε_{it} is independent and identically distributed with $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$.
(ii) For any i and t, V_{it} is independent and identically distributed with $V_{it} \sim N(0, \sigma_v^2 I_k)$.

Assumption 2
The latent factors F_t are distributed independently of $\varepsilon_{it'}$ and $V_{it'}$ for all i, t and t'.

Assumption 3 *The factor loadings γ_i have the following discrete structure*

$$\gamma_i = \overline{\gamma}_{G_j} \qquad (\textit{for all } i \in G_j)$$

for any (i, j) with $1 \leq i \leq N$ and $1 \leq j \leq J_N$. Note that $\mathcal{G} = \{G_1, \ldots, G_{J_N}\}$ is a partition of $\{1, \ldots, N\}$, where J_N is unknown. Let us also suppose that N is very large but J_N is finite ($J_N \ll N$).

Assumption 4
(i) For any $i \in G_j$ with $1 \leq j \leq J_N$, the factor loadings Γ_i are assumed to have the following structure:

$$\Gamma_i^{(l)} = \overline{\Gamma}_{G_j}^{(l)} + \xi_i^{(l)} \ , \quad \xi_i^{(l)} \sim iid\mathcal{L}(0, \Omega_\xi^{(l)})$$

where $1 \leq l \leq k$ and $\Omega_\xi^{(j)}$ is a symmetric and non-negative definite matrix. It is also assumed that ξ_i is distributed independently and identically across i, and of $\varepsilon_{i't}$, $V_{i't}$ and F_t for all i, i' and t.
(ii) Let us also assume

$$\overline{\Gamma}_{G_j} = \overline{\Gamma}$$

for any j with $1 \leq j \leq J_N$.

Assumption 5
(i) $(\log N)/T = o\,(1)$
(ii) $T/N = o\,(1)$

3.2.7 Comment on Assumptions 3 and 4

Assumptions 3 and 4 impose a discrete grouped structure in the factor loadings γ_i and Γ_i, respectively. To understand an intuition of such a structure, it is useful to consider multi-country data. In multi-country data, there could be country (individual), regional (group), and global (common) variations. In this case, the latent factors (common factors) represent global variation, whereas the group structure of the factor loadings generates regional variation. There can be two possible criticisms on Assumptions 3 and 4. First, Assumption 3 imposes only a discrete structure in the loadings γ_i:

$$\gamma_i = \overline{\gamma}_{G_j} \tag{3.13}$$

Note that Eq. (3.13) does not include an individual-specific element corresponding to ξ_i in the loadings Γ_i (see Assumption 4). Second, Assumption 4(ii) is also a restrictive assumption. As for extension to a more realistic setup, it will be discussed in Section 3.5.

3.3 Key properties of the estimators

3.3.1 Lemmas for CCEP and oracle estimators

The lemmas for CCEP and oracle CCE estimators are provided below.

Lemma 6 **[CCEP estimator]** *Suppose Assumption 1 and 2 hold. Then,*

$$\text{(i)} \quad \sqrt{NT}\cdot(\widehat{\beta}_{CCE}-\widehat{\beta}) = \left(\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}^{(p)}\eta_{it}\right)$$

$$\text{(ii)} \quad \lim_{N,T\to\infty}\left[\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right] = \sigma_v^2 I_k + \Delta_{\Sigma}^{(p)}$$

Note also that under some unrestrictive assumptions and approximations, $\Delta_{\Sigma}^{(p)} = o\left(\frac{1}{T}\right) + o\left(\frac{1}{N}\right)$ (for detail, see Section 3.7 and appendix 3.8.4).

Proof. The lemma is proven in 3.7.2, where the sublemmas (i) and (ii) are proven by Lemmas 13 and 12, respectively. □

Lemma 7 **[Oracle CCE estimator]** *Suppose Assumptions 1, 2 and 4 hold. Then,*

$$\text{(i)} \quad \sqrt{NT}\cdot(\widehat{\beta}_{CCE}^{(OR)}-\widehat{\beta}) = \left(\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}\widetilde{X}_{it}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}\eta_{it}\right)$$

$$\text{(ii)} \quad \lim_{N\to\infty}\left[\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}\widetilde{X}_{it}^{\prime}\right] = \Sigma_X^{(OR)} + \sigma_v^2 I_k + \Delta_{\Sigma}^{(OR)}$$

where $\Sigma_X^{(OR)} = \lim_{N\to\infty}\frac{1}{N}\cdot\sum_{i=1}^{N}(\Gamma_i-\overline{\Gamma})^{\prime}(\Gamma_i-\overline{\Gamma}) = Var(\Gamma_i^{\prime})$. Note also that under some unrestrictive assumptions and approximations, $\Delta_{\Sigma}^{(OR)} = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$ (for detail, see Section 3.7 and appendix 3.8.4).

Proof. The lemma is proven in 3.7.3, where the sublemmas (i) and (ii) are proven by Lemmas 17 and 16, respectively. □

For simplicity of the discussion, let us suppose $\eta_{it} \simeq \varepsilon_{it}$ and $\varepsilon_{it} \perp (\widetilde{X}_{it}^{(p)}, \widetilde{X}_{it})$. In this case, the variance of the estimators is obtained as

$$\begin{aligned} Var(\widehat{\beta}_{CCE}) &= E\left[\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\right] \simeq \sigma_{\varepsilon}^2\cdot\left(\sigma_v^2\cdot I_k\right)^{-1} \\ Var(\widehat{\beta}_{CCE}^{(OR)}) &= E\left[\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}\widetilde{X}_{it}^{\prime}\right)^{-1}\right] \simeq \sigma_{\varepsilon}^2\cdot\left(Var(\Gamma_i^{\prime})+\sigma_v^2\cdot I_k\right)^{-1} \end{aligned} \tag{3.14}$$

where $Var(\Gamma_i^{\prime}) = E[(\Gamma_i - \overline{\Gamma})^{\prime}(\Gamma_i - \overline{\Gamma})]$. To give a rough image to Eq. (3.14), let us recall the definition of $\widetilde{X}_{it}^{(p)}$ and $\widetilde{X}_{it}$:

$$\widetilde{X}_{it}^{(p)} = X_{it} - \left(\sum_{s=1}^{T} X_{is}\overline{Z}_s^{\prime}\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s^{\prime}\right)^{-1}\overline{Z}_t \tag{3.15}$$

$$\widetilde{X}_{it} = X_{it} - \frac{1}{N_{g(i)}}\cdot\left(\sum_{s=1}^{T}\sum_{i^{\prime}\in g(i)} X_{i^{\prime}s}\overline{Z}_s^{\prime}\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s^{\prime}\right)^{-1}\overline{Z}_t \tag{3.16}$$

Both $\widetilde{X}_{it}^{(p)}$ and $\widetilde{X}_{it}$ are obtained by projecting X_{it} to the orthogonal space to $\overline{Z}_t$, where the orthogonality is defined by[3]

$$\frac{1}{T}\cdot\sum_{t=1}^{T}\widetilde{X}_{it}^{(p)}\overline{Z}_t^{\prime} = 0 \tag{3.17}$$

$$\frac{1}{N_{g(i)}T}\cdot\sum_{i^{\prime}\in g(i)}\sum_{t=1}^{T}\widetilde{X}_{i^{\prime}t}\overline{Z}_t^{\prime} = 0 \tag{3.18}$$

In this framework, $\widetilde{X}_{it}^{(p)}$ and $\widetilde{X}_{it}$ are approximated by[4]:

$$\begin{cases} \widetilde{X}_{it}^{(p)} \simeq V_{it} \\ \widetilde{X}_{it} \simeq (\Gamma_i^{\prime} - \overline{\Gamma}^{(g(i))\prime})F_t + V_{it} = \xi_i^{\prime}F_t + V_{it} \end{cases} \tag{3.19}$$

$\widetilde{X}_{it}^{(p)}$ and $\widetilde{X}_{it}$ can be interpreted as follows. First, the oracle estimator is an estimator based on true grouping of the factor structure (interactive fixed effects). When the group structure is identified, the projec-

[3]Eqs. (3.17) and (3.18) are proven in Section 3.7 (see Lemmas 11 and 15).

[4]As a rough approximation, let us suppose $V_{it} \perp \overline{Z}_t$, and in this case, it is easy to see

$$\frac{1}{T}\cdot\sum_{t=1}^{T}V_{it}\overline{Z}_t^{\prime} = 0 \quad , \quad \frac{1}{N_{g(i)}T}\cdot\sum_{i^{\prime}\in g(i)}\sum_{t=1}^{T}(\xi_{i^{\prime}}^{\prime}F_t + V_{it})\overline{Z}_t^{\prime} = 0$$

which correspond to Eqs. (3.17) and (3.18).

tion (3.16) is conducted by using the information of all members in the group $g(i)$. Therefore, the obtained $\widetilde{X}_{it}$ includes not only the idiosyncratic disturbances ($= V_{it}$) but also the idiosyncratic responses to the common factors ($= \xi_i' F_t$), where the latter component is extracted as a deviation from the mean response to the common factors at the group level. On the contrary, the CCEP estimator ignores such a group structure. The projection (3.15) is based on the information of the member i only, and as a consequence, the obtained $\widetilde{X}_{it}^{(p)}$ just includes the disturbances V_{it}. Here, Assumption 4(ii) gives $\overline{\Gamma}^{(g(i))} = \overline{\Gamma}$. Then, Eq. (3.19) yields

$$\begin{aligned} Var(\widehat{\beta}_{CCE}) &= \sigma^2 \cdot E\left[(\widetilde{X}_{it}^{(p)} \widetilde{X}_{it}^{(p)\prime})^{-1}\right] \simeq \sigma_\varepsilon^2 \cdot (\sigma_v^2 \cdot I_k)^{-1} \\ Var(\widehat{\beta}_{CCE}^{(OR)}) &= \sigma^2 \cdot E\left[(\widetilde{X}_{it} \widetilde{X}_{it}')^{-1}\right] \simeq \sigma_\varepsilon^2 \cdot \left(Var(\Gamma_i') + \sigma_v^2 \cdot I_k\right)^{-1} \end{aligned} \tag{3.20}$$

where $Var(\Gamma_i') = E[(\Gamma_i - \overline{\Gamma})'(\Gamma_i - \overline{\Gamma})]$.

The above discussion clearly indicates the benefit of group identification in terms of estimation efficiency for β. Note that the benefit is brought by the individual-specific component of the covariates X_{it} in their response to the common factors (i.e. $\xi_i' F_t$).

3.3.2 Discussion on SRC-LASSO estimator

In what follows, we provide a rough discussion on whether the SRC-LASSO estimator owns oracle property.

Framework for a rough discussion For a simplified discussion, let us consider the following DGP:

$$\begin{aligned} y_{it} &= \alpha_i Z_t + \beta X_{it} + \eta_{it} \\ &= \alpha_i Z_t + \beta X_{it} + \varepsilon_{it} - \alpha_i U_t \end{aligned}$$

where $dim(Z_t) = dim(X_{it}) = 1$. As well as our baseline DGP, Z_t and X_{it} is a function of the factors F_t

$$Z_t = CF_t + U_t \ , \quad X_{it} = \Gamma_i F_t + V_{it}$$

where $dim(F_t) = 1$ is also assumed. As for γ_i,

$$\alpha_i = C(CC)^{-1}\gamma_i \implies \gamma_i = C\alpha_i \tag{3.21}$$

Note also that

$$\alpha_i = \sum_{j=1}^{J_N} \alpha^{(G_j)} \cdot \mathbb{I}(i \in G_j) \ , \quad \gamma_i = \sum_{j=1}^{J_N} \gamma^{(G_j)} \cdot \mathbb{I}(i \in G_j)$$

where $\gamma^{(G_j)} = C\alpha^{(G_j)}$. In this framework, the least squares objective function is

$$\begin{aligned} \widehat{Q}^2(a,b) &= \frac{1}{NT} \cdot \sum_{i,t} (y_{it} - a_i Z_t - bX_{it})^2 \\ &= \frac{1}{NT} \cdot \sum_{i,t} \left[(\alpha_i - a_i)Z_t + (\beta - b)X_{it} + \varepsilon_{it} - \alpha_i U_t\right]^2 \end{aligned}$$

Then, the following estimator can be defined

$$(\widehat{\alpha}_{lasso}, \widehat{\beta}_{lasso}) = arg \min_{a,b} \left[\widehat{Q}(a,b) + P_\lambda(a)\right]$$

where $P_\lambda(a) = \widehat{\lambda} \cdot \sum_{i,k} \widehat{w}_{ik} \cdot |a_i - a_k|$. Moreover, set $\boldsymbol{a} = \boldsymbol{m} + \boldsymbol{d}$ where the entries m_i of $\boldsymbol{m}$ are constant across each G_j and $\sum_{i=1}^{n} d_i \mathbb{I}\,(i \in G_j) = 0$, and let $\boldsymbol{M}$ and $\boldsymbol{D}$ be the corresponding sets of $\boldsymbol{m}$'s and $\boldsymbol{d}$'s. Using these notations, an oracle estimator can be also defined:

$$(\widehat{\mu}_{lasso}, \widehat{\beta}^{(OR)}) = arg \min_{m \in M, b} \left[\widehat{Q}(m) + P_\lambda(m)\right]$$

For the above two minimization problems, let us suppose the case of $b \simeq \beta$, which represents that b is sufficiently close to β and $\widehat{Q}$ is no longer a function of b[5]. In such a situation, the following minimizers can be also defined.

$$(\widehat{\alpha}_{lasso})_{b \simeq \beta} = arg \min_{a} \left[\widehat{Q}(a, b \simeq \beta) + P_\lambda(a)\right] \tag{3.22}$$

$$(\widehat{\mu}_{lasso})_{b \simeq \beta} = arg \min_{m \in M} \left[\widehat{Q}(m, b \simeq \beta) + P_\lambda(m)\right] \tag{3.23}$$

[5]This is a very rough concept, but we do not provide a clear definition to it throughout this discussion.

Here, by setting $\gamma_i = C\alpha_i$ (see Eq. (3.21)) and $c_i = Ca_i$, let us introduce the following functions:

$$\begin{aligned} \widehat{Q}^2_{ref}(c) &\equiv \frac{1}{N} \cdot \sum_{i=1}^{N} (c_i - \gamma_i)^2 - \frac{2}{N} \cdot \sum_{i=1}^{N} (c_i - \gamma_i) \cdot \overline{(\varepsilon_i F)} \\ P_{\lambda^{(c)}}(c) &\equiv \widehat{\lambda}^{(c)} \cdot \sum_{i,k} \widehat{w}_{ik} \cdot |c_i - c_k| \end{aligned} \tag{3.24}$$

where $\lambda^{(c)} \equiv C^{-1}\lambda$, and $\overline{(\varepsilon_i F)} \equiv \frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{it} F_t$. Using these functions, the following minimizer can be defined:

$$\widehat{\gamma}^{(ref)}_{lasso} \equiv arg \min_{c} \left[\widehat{Q}_{ref}(c) + P_{\lambda^{(c)}}(c) \right] \tag{3.25}$$

Moreover, set $\boldsymbol{c} = \boldsymbol{m^{(c)}} + \boldsymbol{d^{(c)}}$ where the entries $m_i^{(c)}$ of $\boldsymbol{m}^{(c)}$ are constant across each G_j and $\sum_{i=1}^{n} d_i^{(c)} \mathbb{I}\,(i \in G_j) = 0$, and let $\boldsymbol{M}^{(c)}$ and $\boldsymbol{D}^{(c)}$ be the corresponding sets of $\boldsymbol{m}^{(c)}$'s and $\boldsymbol{d}^{(c)}$'s. Then, the following oracle estimator can be also defined:

$$\widehat{\omega}^{(ref)}_{lasso} \equiv arg \min_{m^{(c)} \in M^{(c)}} \left[\widehat{Q}_{ref}(m^{(c)}) + P_{\lambda^{(c)}}(m^{(c)}) \right] \tag{3.26}$$

<u>A lemma</u> As discussed in 3.7.4, the estimators (3.22)-(3.23) and (3.25)-(3.26) have the following relations[6]:

$$\begin{cases} (\widehat{\alpha}_{lasso})_{b \simeq \beta} & \simeq \; C^{-1} \widehat{\gamma}^{(ref)}_{lasso} \\ (\widehat{\mu}_{lasso})_{b \simeq \beta} & \simeq \; C^{-1} \widehat{\omega}^{(ref)}_{lasso} \end{cases} \tag{3.27}$$

where the approximations hold under $\frac{T}{N} = o(1)$ (i.e. Assumption 5(ii)). Then, the following lemma is obtained.

Lemma 8 *Suppose Assumptions 1(i), 2, 3 and 5(i) hold. Assume there is a* $\delta > 0$ *such that*

$$\widehat{\lambda}^{(c)} \min_{j} \min_{(i,k) \in G_j \times G_j} \{N_j \widehat{w}_{ik}\} \geq \frac{1 + \delta + o_{\mathbb{P}}(1)}{N} \sqrt{\frac{2 \ln N}{T}}. \tag{3.28}$$

For simplicity of the discussion, assume also that $\widehat{\gamma}^{(ref)}_{lasso}$ *exists in a space with finite size* Ω_c ($\subset R^N$)*. Then,* $\widehat{\gamma}^{(ref)}_{lasso} = \widehat{\omega}^{(ref)}_{lasso}$ *with a probability tending to 1.*

[6] For the derivation of Eq. (3.27), see 3.7.4.

Proof. The lemma is proven in 3.7.5. (Note that this lemma is proven by Proposition 24.) $\square$

Lemma 8 shows that the estimator $\widehat{\gamma}^{(ref)}_{lasso}$ owns oracle property, and Eq. (3.27) subsequently suggests that the baseline objective function (i.e. $\widehat{Q}(a,b)+P_\lambda(a)$) should also give oracle property to our SRC-LASSO estimator in the parameter region around $b=\beta$[7]. Note also that Eq. (3.28) does not include the parameter σ_ε. This fact indicates that the penalty level can be chosen without the knowledge of σ_ε[8].

3.4 Monte Carlo experiments

This section is devoted to conducting Monte Carlo (MC) experiments. Based on the implications from the previous section, the main interests are whether the small sample properties of the SRC-LASSO estimator are superior to those of the existing estimator (CCEP estimator) which does not deal with group identification, and whether they are comparable to the properties of the oracle CCE estimator.

3.4.1 DGP

The data generating process (DGP) is defined as follows. Define

$$\begin{aligned} y_{it} &= \beta_1 X_{it,1} + \beta_2 X_{it,2} + \gamma_{i1} F_{1t} + \gamma_{i2} F_{2t} + \varepsilon_{it} \\ X_{it,j} &= \Gamma_{i1,j} F_{1t} + \Gamma_{i3,j} F_{3t} + V_{it,j} \end{aligned} \tag{3.29}$$

for $i=1,2,...,N$, $j=1,2$, and $t=1,2,...,T$. The parameters of interest are β_1 and β_2. The latent factors F_{kt} $(k=1,2,3)$ are generated by AR(1) process where the persistency is set equal to 0.5 ($\rho_f=0.5$). The factor loadings γ_i are assumed to have three discrete groups ($J_N=3$):

$$\gamma_i = (\gamma_{i1},\gamma_{i2})' = \begin{cases} (\widetilde{\gamma}^{(G1)}_{i1},\widetilde{\gamma}^{(G1)}_{i2})' & (i\in G_1) \\ (\widetilde{\gamma}^{(G2)}_{i1},\widetilde{\gamma}^{(G2)}_{i2})' & (i\in G_2) \\ (\widetilde{\gamma}^{(G3)}_{i1},\widetilde{\gamma}^{(G3)}_{i2})' & (i\in G_3) \end{cases}$$

[7]It should be emphasized that the validity of this argument is not exactly guaranteed, because the discussion throughout this subsection is based on approximations and assumptions which are not necessarily relevant. In particular, the assumption of $dim(Z_t)=1$ is not consistent with the model setup we introduced in Section 3.2.

[8]As mentioned in Section 3.1, this is one of the attractive features of the square-root Lasso approach.

Table 3.1: Parameter setup in the DGP

parameters	set value
β	$\beta_1 = 1.0$, $\beta_2 = 1.0$
γ_i	$\gamma_i = \begin{cases} (2.0, -0.5)' & (i \in G_1) \\ (1.5, 0.0)' & (i \in G_2) \\ (1.0, 1.0)' & (i \in G_3) \end{cases}$
μ_Γ	5.0
σ_Γ	2.5
σ_ε	1.0
σ_v	1.0

where the groups G_1, G_2 and G_3 are defined by

$$\begin{array}{ll} G_1 = \{1, ..., N_1\} & (N_1 = 0.3 \times N) \\ G_2 = \{N_1 + 1, ..., N_1 + N_2\} & (N_2 = 0.4 \times N) \\ G_3 = \{N_1 + N_2 + 1, ..., N_1 + N_2 + N_3 (= N)\} & (N_3 = 0.3 \times N) \end{array}$$

As for the loadings Γ_i in the X_{it} equation,

$$\begin{pmatrix} \Gamma_{i1,1} & \Gamma_{i2,1} & \Gamma_{i3,1} \\ \Gamma_{i1,2} & \Gamma_{i2,2} & \Gamma_{i3,2} \end{pmatrix} \sim \begin{pmatrix} iidN(\mu_\Gamma, \sigma_\Gamma^2) & 0 & iidN(0, \sigma_\Gamma^2) \\ iidN(0, \sigma_\Gamma^2) & 0 & iidN(\mu_\Gamma, \sigma_\Gamma^2) \end{pmatrix}$$

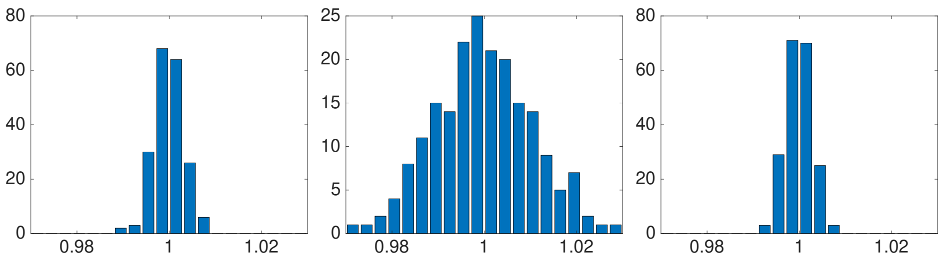
The individual-specific disturbances ε_{it} and $V_{it,j}$ are generated by

$$\begin{array}{rcl} \varepsilon_{it} & \sim & iidN(0, \sigma_\varepsilon^2) \\ V_{it,j} & \sim & iidN(0, \sigma_v^2) \end{array}$$

The free parameters in the DGP are the following: β, γ_i, μ_Γ, σ_Γ, σ_ε, and σ_v. The set values for these parameters are summarized in Table 3.1. It should be also noted that the latent factors (F_t) and the factor loadings (γ_i and Γ_i) are defined so that they satisfy the identification restrictions in Eq. (3.2). For detail, see appendix 3.8.1.

3.4.2 Results

Using the above DGP, we investigate the small sample properties of the CCEP, oracle CCE, and SRC-LASSO estimators. The experiments are replicated 100 times with $(N, T) = (100, 100)$. In examining the Lasso estimator, the penalty parameter λ is optimized by using BIC (Bayesian in-



(a) Lasso (b) CCEP (c) oracle CCE

Figure 3.1: Distributions of $\widehat{\beta}$ across 100 MC experiments (in the baseline case of parameter setup)

Table 3.2: RMSE of $\widehat{\beta}$ and $N(\widehat{J}_N = 3)$ (in the baseline case of parameter setup)

RMSE			$N(\widehat{J}_N = 3)$
$\widehat{\beta}_{lasso}$	$\widehat{\beta}_{CCE}$	$\widehat{\beta}^{(OR)}_{CCE}$	
3.03e-03	1.10e-02	2.80e-03	95

formation criterion)[9]. To solve the minimization problem with ℓ_1-penalty, we use a MATLAB function *fmincon*.

Figure 3.1 shows the distribution of $\widehat{\beta}$ across the 100 experiments, and the obtained RMSE (root mean squared error) is summarized in Table 3.2. As indicated in the figure and table, the Lasso estimator clearly improves RMSE of the CCEP estimator, and the improved RMSE is comparable to that of the oracle CCE estimator. To examine the group identification performance, we also examine the measure $N(\widehat{J}_N = 3)$, which represents the number of experiments where the Lasso estimator identifies three groups in the factor structure (i.e. $\widehat{J}_N = 3$). In Table 3.2, it is revealed that the group identification is successful in most experiments, indicating that the estimator exhibits a good performance in terms of not only RMSE but also group identification. On the whole, the obtained results suggest that the Lasso estimator behaves as if it knows the true grouping of the factor structure, and in this sense, the estimator seems to enjoy oracle property.

To see how the simulation results change depending on the model parameter's setup, a further exercise is conducted. As one trial, Table 3.3 examines several different cases of σ_ε with the other parameters

[9]For detail, see appendix 3.8.2.

set to be their baseline value. Similar exercises are also performed in Tables 3.4-3.6, where σ_v, μ_Γ and σ_Γ are changed, respectively. The parameter dependence of RMSE is also depicted in Figure 3.2.

As to RMSE of the CCEP and oracle CCE estimators, the finding is consistent with Eq. (3.20). As indicated in Tables 3.3-3.5, the RMSE of both estimators is roughly proportional to σ_ε, inversely proportional to σ_v, and shows little dependence on μ_Γ[10]. In Table 3.6, the oracle estimator exhibits a negative dependence on σ_Γ, whereas the behavior of the CCEP estimator is hardly affected by this parameter. Note also that when $\sigma_\Gamma = 0$, the oracle estimator hardly improves RMSE of the CCEP estimator. As already mentioned in 3.3.1, the estimator obtains a benefit of group identification only if the covariates X_{it} own an individual-specific component in their response to the common factors (i.e. $\xi_i' F_t$).

Tables 3.3-3.6 and Figure 3.2 also report the performance of the Lasso estimator. In most cases of the parameter setup, the estimator exhibits an adequate performance of group identification (see the result for $N(\widehat{J}_N = 3)$), and in such a case, the obtained RMSE tends to be comparable to that of the oracle CCE. In terms of RMSE, the Lasso estimator exhibits a much better behavior than the CCEP estimator in a wide range of the model parameter's space at the DGP level.

It should be also mentioned that there are some cases (i.e. the cases of large σ_ε or $\mu_\Gamma = 0$) where the Lasso estimator exhibits a worse performance than the oracle CCE estimator. This issue is caused by the endogenous problem described in 3.2.1. A further discussion is provided in the next subsection.

3.4.3 Impact of endogenous problem

As described in 3.2.1, the regression equation (3.5) owns an endogenous problem coming from correlation between $\overline{Z}_t$ and η_{it}. The definitions of $\overline{Z}_t$ (see (3.3) and (3.4)) and η_{it} (see (3.6)) suggest that the regression model is likely to suffer from this problem in at least the following two cases: when the variance of ε_{it} is large; and when the scale of the factor loadings Γ is small. Figure 3.2 exhibits that RMSE of the Lasso estima-

[10]In Table 3.5, RMSE of the two estimators is slightly deteriorated when $\mu_\Gamma = 0$. This issue is related to the endogenous problem described in 3.2.1. More detailed discussion will be provided in the next subsection.

Table 3.3: RMSE of $\widehat{\beta}$ and $N(\widehat{J}_N = 3)$ in several cases of σ_ε

σ_ε	RMSE			$N(\widehat{J}_N = 3)$
	$\widehat{\beta}_{lasso}$	$\widehat{\beta}_{CCE}$	$\widehat{\beta}^{(OR)}_{CCE}$	
0.1	3.31e-04	1.80e-03	3.05e-04	97
0.5	1.49e-03	5.70e-03	1.41e-03	100
1.0 (baseline)	3.03e-03	1.10e-02	2.80e-03	95
2.0	1.06e-02	2.14e-02	5.59e-03	11
5.0	3.19e-02	5.27e-02	1.40e-02	1

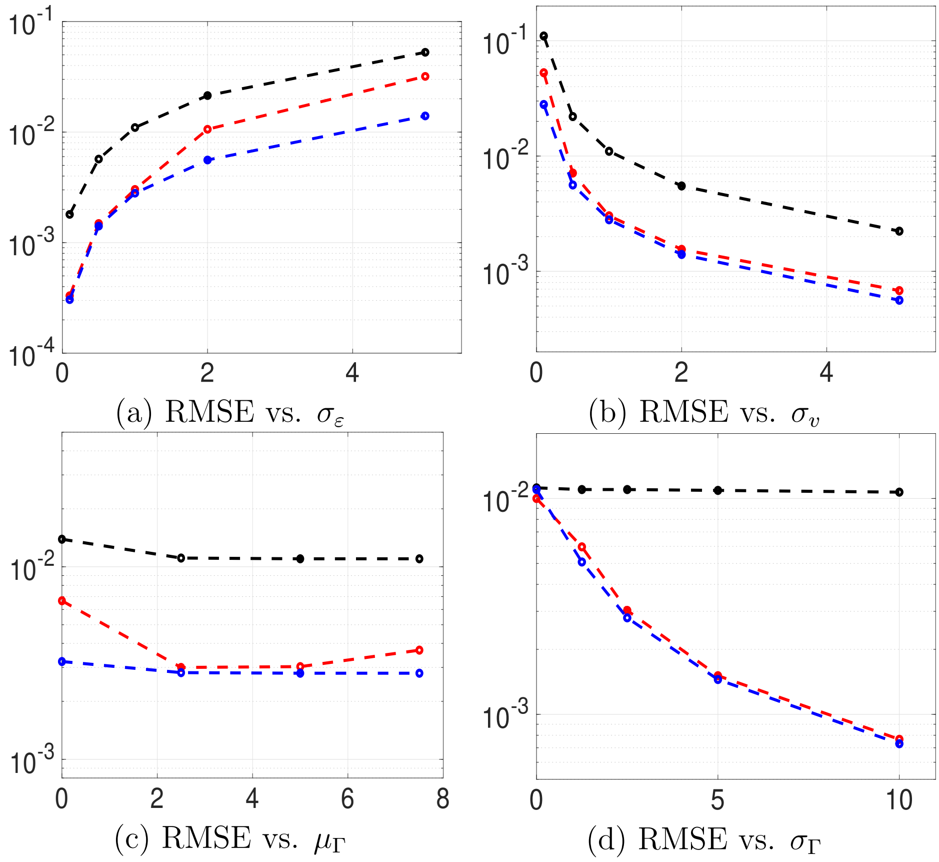
Table 3.4: RMSE of $\widehat{\beta}$ and $N(\widehat{J}_N = 3)$ in several cases of σ_v

σ_v	RMSE			$N(\widehat{J}_N = 3)$
	$\widehat{\beta}_{lasso}$	$\widehat{\beta}_{CCE}$	$\widehat{\beta}^{(OR)}_{CCE}$	
0.1	5.28e-02	1.10e-01	2.80e-02	88
0.5	7.12e-03	2.20e-02	5.61e-03	94
1.0 (baseline)	3.03e-03	1.10e-02	2.80e-03	95
2.0	1.55e-03	5.50e-03	1.40e-03	95
5.0	6.80e-04	2.23e-03	5.61e-04	98

tor is clearly worse than that of the oracle CCE estimator in each case of large σ_ε and $\mu_\Gamma = 0$. This indicates that the endogenous problem deteriorates the performance of the Lasso estimator so severely that the estimator no longer enjoys oracle property.

It should be recalled that the endogenous problem is faced by both of the oracle CCE and Lasso estimators. Nevertheless, the latter one would be affected more seriously, to the degree that it needs to estimate the group structure. When $\overline{Z}_t$ and η_{it} are correlated to a non-negligible extent, this correlation adversely affects the estimation of the factor loadings α. In the case of the Lasso estimator, the correlation can cause the group misidentification, which gives a further adverse effect to the estimation of β. As a consequence, the Lasso estimator suffers from the endogenous problem more severely than the oracle estimator.

In Tables 3.3 and 3.5, the result for $N(\widehat{J}_N = 3)$ show that the Lasso estimator fails to adequately identify the group structure in both cases of large σ_ε and small $|\mu_\Gamma|$. This observation clearly supports the above argument.



(a) RMSE vs. σ_ε

(b) RMSE vs. σ_v

(c) RMSE vs. μ_Γ

(d) RMSE vs. σ_Γ

Figure 3.2: RMSE vs. model parameters: Black, blue, and red lines represent RMSE of CCEP, oracle CCE, and SRC-Lasso estimators, respectively.

Table 3.5: RMSE of $\widehat{\beta}$ and $N(\widehat{J}_N = 3)$ in several cases of μ_Γ

μ_Γ	RMSE			$N(\widehat{J}_N = 3)$
	$\widehat{\beta}_{lasso}$	$\widehat{\beta}_{CCE}$	$\widehat{\beta}^{(OR)}_{CCE}$	
0.0	6.66e-03	1.39e-02	3.22e-03	0
2.5	3.00e-03	1.11e-02	2.82e-03	84
5.0 (baseline)	3.03e-03	1.10e-02	2.80e-03	95
7.5	3.69e-03	1.10e-02	2.80e-03	95

Table 3.6: RMSE of $\widehat{\beta}$ and $N(\widehat{J}_N = 3)$ in several cases of σ_Γ

σ_Γ	RMSE			$N(\widehat{J}_N = 3)$
	$\widehat{\beta}_{lasso}$	$\widehat{\beta}_{CCE}$	$\widehat{\beta}^{(OR)}_{CCE}$	
0.0	1.00e-02	1.12e-02	1.10e-02	93
1.25	5.97e-03	1.10e-02	5.08e-03	94
2.5 (baseline)	3.03e-03	1.10e-02	2.80e-03	95
5.0	1.51e-03	1.09e-02	1.45e-03	94
10.0	7.66e-04	1.07e-02	7.31e-04	95

3.5 Comments on model extension

As described in 3.2.7, there can be two criticisms of our model setup. In what follows, we discuss what kind of model extension would be possible to overcome such criticisms.

3.5.1 Removal of Assumption 4(ii)

In this study, Assumption 4(ii) is introduced for simplicity of the discussion in 3.3.1 and the setup of MC experiments in Section 3.4. Actually, it would not be a hard work to remove this assumption. The variance of the oracle estimator in Eq. (3.20) can be generalized by

$$Var(\widehat{\beta}^{(OR)}_{CCE}) \simeq \sigma^2_\varepsilon \cdot \left(Var(\xi'_i) + \sigma^2_v \cdot I_k\right)^{-1} \quad \text{with } Var(\xi'_i) = E[\xi'_i \xi_i]$$

which holds even after Assumption 4(ii) is removed. The implication of this equation is the same as the one given in 3.3.1: the oracle estimator should be more efficient than an estimator which does not deal with group identification, and this benefit is brought by the individual-specific component of the covariates X_{it} in their response to the common factors (i.e. $\xi_i' F_t$).

3.5.2 A more realistic setup of the loadings γ_i

As described in 3.2.7, Assumption 3 imposes only a discrete structure in the loadings γ_i and ignores an individual-specific element. As a more realistic setup, it is possible to consider the following structure:

$$\gamma_i = \overline{\gamma}_{G_j} + \zeta_i$$

where ζ_i represents an individual-specific element. In the literature on discrete estimators, numerous papers study the properties of their method under the assumption that individual heterogeneity is discrete in the population. An exception is the work by Bonhomme et al. (2017), who examine the properties of the grouped fixed-effects method when individual heterogeneity is not necessarily discrete at the DGP level. An extension of their methodology might give a useful discussion to our work, though it is beyond the scope of this study.

3.6 Conclusion

This study proposes a novel estimation methodology of panel data models with unobserved group multifactor structures. The parameters of interest are the regressions coefficients of the observed covariates. Within the Pesaran (2006)'s common correlated effects (CCE) regression framework, we propose to estimate the model by minimizing the square root of the sum of least squared errors with a shrinkage penalty (Square-root Lasso approach). By introducing ℓ_1-constraint of the pair-wise difference of the factor loadings, the latent group structures including unknown group memberships are identified. In this study, the properties of the square-root Lasso estimator are investigated in comparison with two other estimators. One is the existing estimator which does not deal with group identification. The second one is an oracle estimator, an infeasible estimator based on the true grouping of the factor structure. Through a series of theoretical discussion, it is shown that the oracle estimator is more efficient than the existing estimator under some simple conditions at the population level. Using a simplified model framework, it is also argued that the Lasso estimator is likely to acquire oracle property. To examine the small sample properties of the Lasso estimator, we subsequently conduct Monte-Carlo (MC) experiments. In a wide range of model parameter's space, the Lasso estimator improves RMSE of the existing estimator to a substantial extent, and it also exhibits a comparable behavior to the oracle estimator. It should be also mentioned that the performance of the Lasso estimator can be deteriorated when an endogenous problem is non-negligible. This problem gives an adverse effect on the group identification performance, due to which the estimator's behavior becomes worse even in comparison with the oracle estimator (i.e. the estimator loses oracle property).

3.7 Proofs section

3.7.1 Derivation of oracle CCE estimator

In this subsection, the oracle CCE estimator (3.10) is derived. For the model (3.9), the SSE is defined by

$$SSE = \sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} (y_{it} - \beta' X_{it} - \alpha'_{G_j} \overline{Z}_t)^2$$

with the following first order condition:

$$\frac{\partial(SSE)}{\partial \alpha_{G_j}} = -2 \cdot \sum_{t=1}^{T} \sum_{i \in G_j} (y_{it} - \beta' X_{it} - \alpha'_{G_j} \overline{Z}_t) \overline{Z}'_t = 0 \tag{3.30}$$

$$\frac{\partial(SSE)}{\partial \beta} = -2 \cdot \sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} (y_{it} - \beta' X_{it} - \alpha'_{G_j} \overline{Z}_t) X'_{it} = 0 \tag{3.31}$$

From (3.30),

$$
\begin{aligned}
\alpha'_{G_j}\left(\sum_{t=1}^{T}\sum_{i\in G_j}\overline{Z}_t\overline{Z}'_t\right) &= \sum_{t=1}^{T}(y_{it}-\beta' X_{it})\overline{Z}'_t \\
\alpha'_{G_j}\left(\sum_{t=1}^{T}\overline{Z}_t\overline{Z}'_t\right) &= \frac{1}{N_j}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}(y_{it}-\beta' X_{it})\overline{Z}'_t \\
\left(\sum_{t=1}^{T}\overline{Z}_t\overline{Z}'_t\right)\alpha_{G_j} &= \sum_{t=1}^{T}\frac{\overline{Z}_t}{N_j}\left(\sum_{i\in G_j}(y_{it}-\beta' X_{it})\right)
\end{aligned}
$$

Therefore,

$$
\alpha_{G_j} = \left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}'_s\right)^{-1}\left[\sum_{s=1}^{T}\frac{\overline{Z}_s}{N_j}\left(\sum_{k\in G_j}(y_{ks}-\beta' X_{ks})\right)\right] \tag{3.32}
$$

From (3.31),

$$
\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}(y_{it}-\beta' X_{it}-\alpha'_{G_j}\overline{Z}_t)X'_{it} = 0
$$

By substituting (3.32) to this,

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}(y_{it}-\beta' X_{it})X'_{it} \\
&\quad -\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}y_{ks}\overline{Z}'_s\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}'_s\right)^{-1}\overline{Z}_t X'_{it} \\
&\quad +\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}\beta' X_{ks}\overline{Z}'_s\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}'_s\right)^{-1}\overline{Z}_t X'_{it} \quad = \quad 0
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\left[y_{it}-\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}y_{ks}\overline{Z}'_s\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}'_s\right)^{-1}\overline{Z}_t\right]X'_{it} \\
&\quad -\beta'\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\left[X_{it}-\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}'_s\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}'_s\right)^{-1}\overline{Z}_t\right]X'_{it} \quad = \quad 0
\end{aligned}
$$

This can be rewritten as

$$
\left(\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\widetilde{y}_{it}X'_{it}\right)-\beta'\left(\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\widetilde{X}_{it}X'_{it}\right)=0 \tag{3.33}
$$

where

$$
\begin{aligned}
\widetilde{y}_{it} &\equiv y_{it} - \frac{1}{N_{g(i)}} \cdot \left(\sum_{s=1}^{T} \sum_{k \in g(i)} y_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \overline{Z}_t \\
\widetilde{X}_{it} &\equiv X_{it} - \frac{1}{N_{g(i)}} \cdot \left(\sum_{s=1}^{T} \sum_{k \in g(i)} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \overline{Z}_t
\end{aligned}
$$

From (3.33),

$$
\widehat{\beta}_{CCE}^{(OR)} = \left(\sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} X_{it} \widetilde{X}_{it}' \right)^{-1} \left(\sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} X_{it} \widetilde{y}_{it} \right) \tag{3.34}
$$

Here, let us define

$$
\begin{aligned}
\Delta_{it} &\equiv X_{it} - \widetilde{X}_{it} \\
&= \frac{1}{N_j} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \overline{Z}_t
\end{aligned}
$$

for any $i \in G_j$. Using this notation, it is easy to show[11]

$$
\begin{aligned}
\sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} \Delta_{it} \widetilde{y}_{it} &= 0 \\
\sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} \Delta_{it} \widetilde{X}_{it}' &= 0
\end{aligned} \tag{3.35}
$$

which yield

$$
\begin{aligned}
\sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} \widetilde{X}_{it} \widetilde{y}_{it} &= \sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} X_{it} \widetilde{y}_{it} - \sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} \Delta_{it} \widetilde{y}_{it} \\
&= \sum_{t=1}^{T} \sum_{j=1}^{J_N} \sum_{i \in G_j} X_{it} \widetilde{y}_{it}
\end{aligned} \tag{3.36}
$$

[11]For the derivation, see appendix 3.8.3.

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\widetilde{X}_{it}\widetilde{X}_{it}' &= \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}X_{it}\widetilde{X}_{it}' - \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\Delta_{it}\widetilde{X}_{it}' \\
&= \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}X_{it}\widetilde{X}_{it}'
\end{aligned}
\tag{3.37}
$$

Using (3.36) and (3.37), the estimator (3.34) can be rewritten as

$$
\widehat{\beta}_{CCE}^{(OR)} = \left(\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\widetilde{X}_{it}\widetilde{y}_{it}\right)
$$

which yields Eq. (3.10).

3.7.2 Lemmas for CCEP estimator

The goal in this subsection is to prove Lemma 6, which is done by Lemmas 12 and 13. (Note that the sublemmas 6(i) and (ii) correspond to Lemmas 13 and 12, respectively.)

Lemma 9

$$
\overline{C}(\overline{C}'\overline{C})^{-1}\overline{C}' = I_m
$$

Proof. From $C = (\overline{\Gamma}\beta + \gamma\ ,\ \overline{\Gamma})$,

$$
C'C = \begin{pmatrix} (\overline{\Gamma}\beta+\overline{\gamma})'(\overline{\Gamma}\beta+\overline{\gamma}) & (\overline{\Gamma}\beta+\overline{\gamma})'\overline{\Gamma} \\ \overline{\Gamma}'(\overline{\Gamma}\beta+\overline{\gamma}) & \overline{\Gamma}'\overline{\Gamma} \end{pmatrix}
$$

$$
C'\overline{\Gamma} = \begin{pmatrix} (\overline{\Gamma}\beta+\overline{\gamma})'\overline{\Gamma} \\ \overline{\Gamma}'\overline{\Gamma} \end{pmatrix}
$$

which yield

$$
(C'C)^{-1}C'\overline{\Gamma} = \begin{pmatrix} 0_{1\times k} \\ I_k \end{pmatrix}
$$

Therefore,

$$
\begin{aligned}
C(C'C)^{-1}C'\overline{\Gamma} = [\overline{\Gamma}\beta + \overline{\gamma}\ \ \overline{\Gamma}] \begin{pmatrix} 0_{1\times k} \\ I_k \end{pmatrix} = \overline{\Gamma} \\
\implies C(C'C)^{-1}C'\overline{\Gamma} = \overline{\Gamma} \\
\implies C(C'C)^{-1}C'\overline{\Gamma}\overline{\Gamma}' = \overline{\Gamma}\overline{\Gamma}' \\
\implies C(C'C)^{-1}C'\overline{\Gamma}\overline{\Gamma}'(\overline{\Gamma}\overline{\Gamma}')^{-1} = \overline{\Gamma}\overline{\Gamma}'(\overline{\Gamma}\overline{\Gamma}')^{-1} \\
\implies C(C'C)^{-1}C' = I_m \qquad \square
\end{aligned}
$$

Lemma 10 *Suppose Assumptions 1 and 2 hold. Then,*
(i) $\widetilde{y}_{it}^{(p)} = y_{it} - \left(\alpha_i + (\overline{C}'\overline{C})^{-1}\overline{C}'\Gamma_i\beta\right)'\overline{Z}_t + \psi_1\overline{Z}_t$
(ii) $\widetilde{X}_{it}^{(p)} = X_{it} - \left((\overline{C}'\overline{C})^{-1}\overline{C}'\Gamma_i\right)'\overline{Z}_t + \psi_2\overline{Z}_t$
(iii) $\widetilde{y}_{it}^{(p)} = \beta'\widetilde{X}_{it}^{(p)} + \eta_{it} + \psi\overline{Z}_t$
Note also that under some unrestrictive assumptions and approximations, $\psi_1, \psi_2, \psi \sim o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$ *(for detail, see appendix 3.8.4).*

Proof.
(i)

$$
\widetilde{y}_{it}^{(p)} = y_{it} - \left(\frac{1}{T}\cdot\sum_{s=1}^{T} y_{is}\overline{Z}_s'\right)\left(\frac{1}{T}\cdot\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t
$$

From the definition of $y_{it}(= \beta'X_{it} + \alpha_i'Z_t + \eta_{it})$ and $Z_t(= C'F_t + U_t)$,

$$
\begin{aligned}
\frac{1}{T}\cdot\sum_{t=1}^{T} y_{it}Z_t' &= \frac{1}{T}\cdot\sum_{t=1}^{T}\left[(\alpha_i'C' + \beta'\Gamma_i')F_t + \alpha_i'U_t + \beta'V_{it} + \eta_{it}\right]\left[F_t'C + U_t'\right] \\
&= \left(\frac{1}{T}\cdot\sum_{t=1}^{T}(\alpha_i'C' + \beta'\Gamma_i')F_tF_t'C'\right) + \psi_1^{(a)} \\
&= \alpha_i'C'C + \beta'\Gamma_i'C + \psi_1^{(a)}
\end{aligned}
$$

where $\psi_1^{(a)} = o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[12]. Furthermore,

$$
\begin{aligned}
\frac{1}{T}\cdot\sum_{t=1}^{T} Z_tZ_t' &= \frac{1}{T}\cdot\sum_{t=1}^{T}(C'F_t + U_t)(F_t'C + U_t') \\
&= \left(\frac{1}{T}\cdot\sum_{t=1}^{T} C'F_tF_t'C\right) + \psi_1^{(b)} \\
&= C'C + \psi_1^{(b)}
\end{aligned}
$$

[12]For the order estimation, see appendix 3.8.4.

where $\psi_1^{(b)} = o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[13]. Therefore,

$$\begin{aligned}
\widetilde{y}_{it}^{(p)} &= y_{it} - (\alpha_i' C'C + \beta'\Gamma_i' C + \psi_1^{(a)})(C'C + \psi_1^{(b)})^{-1} Z_t \\
&= y_{it} - \alpha_i' Z_t - \beta'\Gamma_i' C (C'C)^{-1} Z_t + \psi_1 Z_t
\end{aligned}$$

where $\psi_1 = o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$. $\square$

(ii)

$$\widetilde{X}_{it}^{(p)} = X_{it} - \left(\frac{1}{T}\cdot\sum_{s=1}^{T} X_{is} Z_s'\right)\left(\frac{1}{T}\cdot\sum_{s=1} Z_s Z_s'\right)^{-1} Z_t$$

From the definition of $X_{it}(= \Gamma_i' F_t + V_{it})$ and $Z_t(= C'F_t + U_t)$,

$$\begin{aligned}
\frac{1}{T}\cdot\sum_{t=1}^{T} X_{it} Z_t' &= \frac{1}{T}\cdot\sum_{t=1}^{T}(\Gamma_i' F_t + V_{it})(F_t' C + U_t') \\
&= \left(\frac{1}{T}\cdot\sum_{t=1}^{T}\Gamma_i' F_t F_t' C\right) + \psi_2^{(a)} \\
&= \Gamma_i' C + \psi_2^{(a)}
\end{aligned}$$

where $\psi_2^{(a)} = o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[14]. Hence,

$$\begin{aligned}
\widetilde{X}_{it}^{(p)} &= X_{it} - (\Gamma_i' C + \psi_2^{(a)})(C'C + \psi_1^{(b)})^{-1} Z_t \\
&= X_{it} - \Gamma_i' C (C'C)^{-1} Z_t + \psi_2 Z_t
\end{aligned}$$

where $\psi_2 = o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$. $\square$

(iii)

$$\begin{aligned}
&\widetilde{y}_{it}^{(p)} - \beta'\widetilde{X}_{it}^{(p)} \\
&\quad = (y_{it} - \alpha_i' Z_t - \beta'\Gamma_i' C (C'C)^{-1} Z_t + \psi_1 Z_t) \\
&\qquad - \beta'(X_{it} - \Gamma_i' C (C'C)^{-1} Z_t + \psi_2 Z_t) \\
&\quad = y_{it} - \beta' X_{it} - \alpha_i' Z_t + \psi Z_t \\
&\quad = \eta_{it} + \psi Z_t
\end{aligned}$$

where $\psi = o\left(\frac{1}{\sqrt{T}}\right) + o\left(\frac{1}{N}\right)$. $\square$

[13] For the order estimation, see appendix 3.8.4.

[14] For the order estimation, see appendix 3.8.4.

Lemma 11 *For any i with $1 \leq i \leq N$,*

$$\sum_{t=1}^{T} \widetilde{y}_{it}^{(p)} \overline{Z}_t' = 0 \ , \quad \sum_{t=1}^{T} \widetilde{X}_{it}^{(p)} \overline{Z}_t' = 0$$

Proof.

$$\begin{aligned}
\sum_{t=1}^{T} \widetilde{y}_{it}^{(p)} \overline{Z}_t' &= \sum_{t=1}^{T} \left[y_{it} - \left(\sum_{s=1}^{T} y_{is} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \overline{Z}_t \right] \overline{Z}_t' \\
&= \left(\sum_{t=1}^{T} y_{it} \overline{Z}_t' \right) - \left(\sum_{s=1}^{T} y_{is} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{t=1}^{T} \overline{Z}_t \overline{Z}_t' \right) \\
&= \left(\sum_{t=1}^{T} y_{it} \overline{Z}_t' \right) - \left(\sum_{s=1}^{T} y_{is} \overline{Z}_s' \right) \\
&= 0
\end{aligned}$$

Likewise, it can be also shown that

$$\sum_{t=1}^{T} \widetilde{X}_{it}^{(p)} \overline{Z}_t' = 0 \qquad \square$$

Lemma 12 *Suppose Assumptions 1 and 2 hold. Then,*

$$\lim_{N,T \to \infty} \frac{1}{NT} \cdot \left(\sum_{i,t} \widetilde{X}_{it}^{(p)} \widetilde{X}_{it}^{(p)\prime} \right) = \sigma_v^2 I + \Delta_{\Sigma}^{(p)}$$

where

$$\Delta_{\Sigma}^{(p)} = o\left(\frac{1}{T}\right) + o\left(\frac{1}{N}\right)$$

under some unrestrictive assumptions and approximations (for detail, see appendix 3.8.4).

Proof.

$$\begin{aligned}
\widetilde{X}_{it}^{(p)} &= X_{it} - \Gamma_i' C (C'C)^{-1} Z_t + \psi_2 Z_t \\
&= \Gamma_i' F_t + V_{it} - \Gamma_i' C (C'C)^{-1} (C' F_t + U_t) + \psi_2 Z_t \\
&= \Gamma_i' \left[I - C (C'C)^{-1} C' \right] F_t + V_{it} - \Gamma_i' C (C'C)^{-1} U_t + \psi_2 Z_t \\
&= V_{it} - \Gamma_i' C (C'C)^{-1} U_t + \psi_2 Z_t
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime} &= \left(\frac{1}{NT}\cdot\sum_{i,t}V_{it}V_{it}'\right)+\Delta_{\Sigma}^{(p)}\\
&= \sigma_v^2 I+\Delta_{\Sigma}^{(p)}
\end{aligned}$$

where $\Delta_{\Sigma}^{(p)} = o\left(\frac{1}{T}\right)+o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[15]. $\square$

Lemma 13 *Suppose Assumptions 1 and 2 hold. Then,*

$$\begin{aligned}
\sqrt{NT}\cdot(\widehat{\beta}_{CCE}-\beta) &= plim\left[\left(\frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\right]\\
&\qquad\left(\frac{1}{\sqrt{NT}}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^{(p)}\eta_{it}\right)
\end{aligned}$$

as $N,T\to\infty$.

Proof. By substituting Lemma 10(iii) to Eq. (3.7),

$$\begin{aligned}
\widehat{\beta}_{CCE} &= \left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}[\widetilde{X}_{it}^{(p)\prime}\beta+\eta_{it}+Z_t'\psi']\right)\\
&= \beta+\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}(\eta_{it}+Z_t'\psi')\right)\\
&= \beta+\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)^{-1}\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}^{(p)}\eta_{it}\right) \qquad \text{(from Lemma 11)}
\end{aligned}$$

Note that Lemma 12 implies

$$\left(\frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^{(p)}\widetilde{X}_{it}^{(p)\prime}\right)\xrightarrow{p}\sigma_v^2 I$$

[15]For the order estimation, see appendix 3.8.4.

as $N, T \to \infty$. Therefore,

$$\sqrt{NT} \cdot (\widehat{\beta}_{CCE} - \beta) = \text{plim} \left[\left(\frac{1}{NT} \cdot \sum_{t=1}^{T} \sum_{i=1}^{N} \widetilde{X}_{it}^{(p)} \widetilde{X}_{it}^{(p)\prime} \right)^{-1} \right] \left(\frac{1}{\sqrt{NT}} \cdot \sum_{t=1}^{T} \sum_{i=1}^{N} \widetilde{X}_{it}^{(p)} \eta_{it} \right)$$

as $N, T \to \infty$. $\square$

3.7.3 Lemmas for oracle CCE estimator

The goal in this subsection is to prove Lemma 7, which is done by Lemmas 16 and 17. (Note that the sublemmas 7(i) and (ii) correspond to Lemmas 17 and 16, respectively.)

Lemma 14 *Suppose Assumptions 1, 2, 3 and 4 hold. Then,*
(i) $\widetilde{y_{it}} = y_{it} - \left(\alpha_i + (\overline{C}'\overline{C})^{-1}\overline{C}'\overline{\Gamma}\beta\right)' \overline{Z}_t + \delta_1 \overline{Z}_t$
(ii) $\widetilde{X}_{it} = X_{it} - \left((\overline{C}'\overline{C})^{-1}\overline{C}'\overline{\Gamma}\right)' \overline{Z}_t + \delta_2 Z_t$
(iii) $\widetilde{y}_{it} = \beta' \widetilde{X}_{it} + \eta_{it} + \delta \overline{Z}_t$
Note also that under some unrestrictive assumptions and approximations, $\delta_1, \delta_2, \delta \sim o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$ *(for detail, see appendix 3.8.4).*

Proof.
(i) For any $i \in G_j$,

$$\widetilde{y}_{it} - \frac{1}{N_j} \cdot \left(\frac{1}{T} \cdot \sum_{s=1}^{T} \sum_{k \in G_j} y_{ks} Z_s' \right) \left(\frac{1}{T} \cdot \sum_{s=1}^{T} Z_s Z_s' \right)^{-1} Z_t$$

Here, from the definition of $y_{it} (= \beta' X_{it} + \alpha_i' Z_t + \eta_{it})$ and $Z_t (= C' F_t + U_t)$, it is easy to show

$$\begin{aligned}
\frac{1}{N_j T} \cdot \sum_{t=1}^{T} \sum_{i \in G_j} y_{it} Z_t' &= \frac{1}{N_j T} \cdot \sum_{t=1}^{T} \sum_{i \in G_j} \left[(\alpha_i' C' + \beta' \Gamma_i') F_t + \alpha_i' U_t + \beta' v_{it} + \eta_{it} \right] \left[F_t' C + U_t' \right] \\
&= \left(\frac{1}{N_j T} \cdot \sum_{t=1}^{T} \sum_{i \in G_j} (\alpha_i' C' + \beta' \Gamma_i') F_t F_t' C' \right) + \delta_1^{(a)} \\
&= \frac{1}{N_j} \cdot \sum_{i \in G_j} (\alpha_i' C' + \beta' \Gamma_i') C + \delta_1^{(a)} \\
&= \alpha_{G_j}' C' C + \beta' \overline{\Gamma}' C + \delta_1^{(a)}
\end{aligned}$$

where $\delta_1^{(a)} = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[16]. Furthermore,

$$\begin{aligned}\frac{1}{T}\cdot\sum_{t=1} Z_t Z_t' &= \frac{1}{T}\cdot\sum_{t=1}^{T}(C'F_t + U_t)(F_t'C' + U_t') \\ &= \left(\frac{1}{T}\cdot\sum_{t=1}^{T} C'F_tF_t'C\right) + \delta_1^{(b)} \\ &= C'C + \delta_1^{(b)}\end{aligned}$$

where $\delta_1^{(b)} = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[17]. Therefore,

$$\begin{aligned}\widetilde{y}_{it} &= y_{it} - (\alpha_{G_j}'C'C + \beta'\overline{\Gamma}C + \delta_1^{(a)})(C'C + \delta_1^{(b)})^{-1}Z_t \\ &= y_{it} - \alpha_{G_j}'Z_t - \beta'\overline{\Gamma}C(C'C)^{-1}Z_t + \delta_1 Z_t \\ &= y_{it} - \alpha_i'Z_t - \beta'\overline{\Gamma}C(C'C)^{-1}Z_t + \delta_1 Z_t\end{aligned}$$

where $\delta_1 = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$. $\square$

(ii) For any $i \in G_j$,

$$\widetilde{X}_{it} = X_{it} - \left(\frac{1}{N_jT}\cdot\sum_{s=1}^{T}\sum_{k\in G_j} X_{is}Z_s'\right)\left(\frac{1}{T}\cdot\sum_{s=1}^{T} Z_sZ_s'\right)^{-1} Z_t$$

Here, from the definition of $X_{it}(=\Gamma_i'F_t + v_{it})$ and $Z_t(= C'F_t + U_t)$,

$$\begin{aligned}\frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j} X_{it}Z_t' &= \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}(\Gamma_i'F_t + v_{it})(F_t'C + U_t') \\ &= \left(\frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\Gamma_i'F_tF_t'C\right) + \delta_2^{(a)} \\ &= \left(\frac{1}{N_j}\cdot\sum_{i\in G_j}\Gamma_i'C\right) + \delta_2^{(a)} \\ &= \overline{\Gamma}'C + \delta_2^{(a)}\end{aligned}$$

[16]For the order estimation, see appendix 3.8.4.
[17]For the order estimation, see appendix 3.8.4

where $\delta_2^{(a)} = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[18]. Therefore,

$$\begin{aligned}
\widetilde{X}_{it} &= X_{it} - (\overline{\Gamma}'C + \delta_2^{(a)})(C'C + \delta_1^{(b)})^{-1}Z_t \\
&= X_{it} - \overline{\Gamma}'C(C'C)^{-1}Z_t + \delta_2 Z_t
\end{aligned}$$

where $\delta_2 = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$. $\square$

(iii)

$$\begin{aligned}
\widetilde{y}_{it} - \beta'\widetilde{X}_{it} &= [y_{it} - \alpha_i' Z_t - \beta'\overline{\Gamma}'C(C'C)^{-1}Z_t + \delta_1 Z_t] \\
&\qquad - \beta'[X_{it} - \overline{\Gamma}'C(C'C)^{-1}Z_t + \delta_2 Z_t] \\
&= y_{it} - \beta' X_{it} - \alpha_i' Z_t + \delta Z_t \\
&= \eta_{it} + \delta Z_t
\end{aligned}$$

where $\delta = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$. $\square$

Lemma 15 *For any j with $1 \leq j \leq J_N$,*

$$\sum_{i\in G_j}\sum_{t=1}^{T} \widetilde{y}_{it}\overline{Z}_t' = 0 \ , \quad \sum_{i\in G_j}\sum_{t=1}^{T} \widetilde{X}_{it}\overline{Z}_t' = 0$$

Proof.

$$\begin{aligned}
\sum_{i\in G_j}\sum_{t=1}^{T} \widetilde{y}_{it}\overline{Z}_t' &= \sum_{i\in G_j}\sum_{t=1}^{T}\left[y_{it} - \frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j} y_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t\right]\overline{Z}_t' \\
&= \left(\sum_{i\in G_j}\sum_{t=1}^{T} y_{it}\overline{Z}_t'\right)' \\
&\qquad - \frac{1}{N_j}\cdot\sum_{i\in G_j}\left(\sum_{s=1}^{T}\sum_{k\in G_j} y_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{t=1}^{T}\overline{Z}_t\overline{Z}_t'\right) \\
&= \left(\sum_{i\in G_j}\sum_{t=1}^{T} y_{it}\overline{Z}_t'\right)' - \left(\sum_{k\in G_j}\sum_{s=1}^{T} y_{ks}\overline{Z}_s'\right)' \\
&= 0
\end{aligned}$$

[18]For the order estimation, see appendix 3.8.4.

Likewise, it can be also shown

$$\sum_{i\in G_j}\sum_{t=1}^{T}\widetilde{X}_{it}\overline{Z}_t' = 0 \qquad \square$$

Lemma 16 *Suppose Assumptions 1, 2, 3 and 4 hold. Then,*

$$\lim_{N\to\infty}\frac{1}{NT}\left(\sum_{i,t}\widetilde{X}_{it}\widetilde{X}_{it}'\right) \;=\; \Sigma_\Gamma + \sigma_v^2\cdot I + \Delta_\Sigma^{(X)}$$

where

$$\Sigma_\Gamma = \lim_{N\to\infty}\frac{1}{N}\cdot\sum_{i=1}^{N}(\Gamma_i' - \overline{\Gamma}')(\Gamma_i - \overline{\Gamma})$$

Note also that

$$\Delta_\Sigma^{(X)} = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$$

under some unrestrictive assumptions and approximations (for detail, see appendix 3.8.4).

Proof.

Using the previous lemma,

$$\begin{aligned}
\widetilde{X}_{it} &= X_{it} - \overline{\Gamma}'C(C'C)^{-1}(C'F_t + U_t) + \delta_2 Z_t \\
&= \Gamma_i'F_t + V_{it} - \overline{\Gamma}'C(C'C)^{-1}(C'F_t + U_t) + \delta_2 Z_t \\
&= \left[\Gamma_i' - \overline{\Gamma}'C(C'C)^{-1}C'\right]F_t + V_{it} - \overline{\Gamma}'C(C'C)^{-1}U_t + \delta_2 Z_t \\
&= \left[\Gamma_i' - \overline{\Gamma}'\right]F_t + V_{it} - \overline{\Gamma}'C(C'C)^{-1}U_t + \delta_2 Z_t
\end{aligned} \tag{3.38}$$

where Lemma 9 is used in the last line. Therefore,

$$\begin{aligned}
\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}\widetilde{X}_{it}' &= \frac{1}{NT}\cdot\sum_{i,t}\left[\Gamma_i' - \overline{\Gamma}'\right]F_tF_t'\left[\Gamma_i - \overline{\Gamma}\right] \\
&\qquad + \left(\frac{1}{NT}\cdot\sum_{i,t}V_{it}V_{it}'\right) + \Delta_\Sigma^{(X)} \\
&= \left(\frac{1}{N}\cdot\sum_{i=1}^{N}\left[\Gamma_i' - \overline{\Gamma}'\right]\left[\Gamma_i - \overline{\Gamma}\right]\right) + \sigma_v^2\cdot I + \Delta_\Sigma^{(X)}
\end{aligned} \tag{3.39}$$

where $\Delta_\Sigma^{(X)} = o\left(\frac{1}{\sqrt{NT}}\right) + o\left(\frac{1}{N}\right)$ under some unrestrictive assumptions and approximations[19]. $\square$

[19]For the order estimation, see appendix 3.8.4.

Lemma 17 *Suppose Assumptions 1, 2, 3 and 4 hold. Then,*

$$\sqrt{NT}\cdot(\widehat{\beta}^{(OR)}_{CCE}-\beta) \;=\; plim\left[\left(\frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\right]\left(\frac{1}{\sqrt{NT}}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\eta_{it}\right)$$

as $N\to\infty$.

Proof. By substituting Lemma 14(iii) to Eq. (3.10),

$$\begin{aligned}
\widehat{\beta}^{(OR)}_{CCE} &= \left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}[\widetilde{X}_{it}'\beta+\eta_{it}+Z_t'\delta']\right)\\
&= \beta+\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}(\eta_{it}+Z_t'\delta')\right)\\
&= \beta+\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\left(\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}\eta_{it}\right) \quad \text{(from Lemma 15)}
\end{aligned}$$

Note that Lemma 16 implies

$$\left(\frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\widetilde{X}_{it}'\right)\xrightarrow{p}\Sigma_\Gamma+\sigma_v^2 I$$

as $N\to\infty$. Therefore,

$$\sqrt{NT}\cdot(\widehat{\beta}^{(OR)}_{CCE}-\beta) \;=\; \text{plim}\left[\left(\frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\widetilde{X}_{it}'\right)^{-1}\right]\left(\frac{1}{\sqrt{NT}}\cdot\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}\eta_{it}\right)$$

as $N\to\infty$. $\square$

3.7.4 Discussion on SRC-LASSO estimator

Derivation of Eq. (3.27) First, since $Z_t=CF_t+U_t$ and $X_{it}=\Gamma_i F_t+V_{it}$,

$$\begin{aligned}
\widehat{Q}^2(a,b) &= \frac{1}{NT}\cdot\sum_{i,t}[(\alpha_i-a_i)(CF_t+U_t)+(\beta-b)(\Gamma_i F_t+V_{it})+\varepsilon_{it}-\alpha_i U_t]^2\\
&= \frac{1}{NT}\cdot\sum_{i,t}\{[C(\alpha_i-a_i)+(\beta-b)\Gamma_i]F_t-a_iU_t+(\beta-b)V_{it}+\varepsilon_{it}\}^2
\end{aligned}$$

When setting $b \simeq \beta$,

$$
\begin{aligned}
\widehat{Q}^2(a, b \simeq \beta) &\simeq \frac{1}{NT} \cdot \sum_{i,t} \left[C(\alpha_i - a_i)F_t - a_i U_t + \varepsilon_{it}\right]^2 \\
&= \left(\frac{1}{NT} \cdot \sum_{i,t} [C(a_i - \alpha_i)F_t]^2 \right) \\
&\quad - \left(\frac{2}{NT} \cdot \sum_{it} C(a_i - \alpha_i)F_t[\varepsilon_{it} - a_i U_t] \right) + (const.) \\
&= \left(\frac{C^2}{N} \cdot \sum_{i=1}^{N} (a_i - \alpha_i)^2 \right) \\
&\quad - \left(\frac{2C}{N} \cdot \sum_{i=1}^{N} (a_i - \alpha_i) \cdot \left[\overline{(\varepsilon_i F)} - a_i \cdot \overline{(FU)} \right] \right) + (const.)
\end{aligned}
\tag{3.40}
$$

where $\overline{(\varepsilon_i F)} = \frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{it} F_t$ and $\overline{(FU)} = \frac{1}{T} \cdot \sum_{t=1}^{T} F_t U_t$. Note also that in the last line, we used $\sum_{t=1}^{T} F_t^2 = T$.

Here, let us utilize Eq. (3.59):

$$\overline{(FU)} = O\left(\frac{1}{\sqrt{NT}} \right)$$

Also, when ε_{it} and F_t are independent,

$$\overline{(\varepsilon_i F)} = O\left(\frac{1}{\sqrt{T}} \right)$$

Therefore, under the condition of $\frac{T}{N} = o(1)$,

$$|\overline{(\varepsilon_i F)}| \gg |\overline{(FU)}|$$

Using this fact, let us approximate Eq. (3.40) as follows:

$$\widehat{Q}^2(a, b \simeq \beta) \simeq \left(\frac{C^2}{N} \cdot \sum_{i=1}^{N} (a_i - \alpha_i)^2 \right) - \left(\frac{2C}{N} \cdot \sum_{i=1}^{N} (a_i - \alpha_i) \cdot \overline{(\varepsilon_i F)} \right) + (const.)$$

Setting $\gamma_i = C\alpha_i$ and $c_i = C a_i$,

$$
\begin{aligned}
\widehat{Q}^2(a, b \simeq \beta) &\simeq \left(\frac{1}{N} \cdot \sum_{i=1}^{N} (c_i - \gamma_i)^2 \right) - \left(\frac{2}{N} \cdot \sum_{i=1}^{N} (c_i - \gamma_i) \cdot \overline{(\varepsilon_i F)} \right) + (const.) \\
&\equiv \widehat{Q}^2_{ref}(c) + (const.)
\end{aligned}
\tag{3.41}
$$

Furthermore, it should be noticed

$$
\begin{aligned}
P_{\lambda}(a) &= \widehat{\lambda} \cdot \sum_{i,k} \widehat{w}_{ik} \cdot |a_i - a_k| \\
&= (C^{-1}\widehat{\lambda}) \cdot \sum_{i,k} \widehat{w}_{ik} \cdot |Ca_i - Ca_k| \\
&= (C^{-1}\widehat{\lambda}) \cdot \sum_{i,k} \widehat{w}_{ik} \cdot |c_i - c_k| \\
&= P_{\lambda^{(c)}}(c)
\end{aligned}
\tag{3.42}
$$

where $\lambda^{(c)} = C^{-1}\lambda$. Therefore, from (3.41), and (3.42), it is obtained

$$
(\widehat{\alpha}_{lasso})_{b\simeq\beta} \simeq C^{-1}\widehat{\gamma}_{lasso} \ , \quad (\widehat{\mu}_{lasso})_{b\simeq\beta} \simeq C^{-1}\widehat{\omega}_{lasso} \qquad \square
$$

3.7.5 Proof of Lemma 8

In what follows, our goal is to prove Lemma 8, which will be done by Proposition 24.

Model To achieve the above goal, let us begin with introducing the following model (DGP):

$$
y_{it} = \gamma_i' F_t + \varepsilon_{it}, \quad \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right).
$$

where $F'F = TI_r$. It is assumed that the γ_i's can take J_N values, where J_N is unknown. It follows that there exists a partition $\mathcal{G} = \{G_1, \ldots, G_{J_N}\}$ of $\{1, \ldots, N\}$ such that, for any element G of the partition $\gamma_i = \gamma_G$ for all i in G. Let N_j be the number of indices i in G_j, $j = 1, \ldots, J_N$.

For the above DGP, we can define the following Lasso-based estimator:

$$
\widehat{\gamma} = arg \min_c \left\{ \widehat{Q}(c) + P_{\lambda}(c) \right\} \tag{3.43}
$$

where

$$
\widehat{Q}^2(c) = \frac{1}{NT} \cdot \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - c_i' F_t)^2
$$

$$
P_{\lambda}(c) = \lambda \cdot \sum_{i=1}^{N} \sum_{k=1}^{N} \widehat{w}_{ik} \cdot |c_i - c_k|
$$

Note also that

$$\begin{aligned}
\widehat{Q}^2(c) &= \frac{1}{NT}\cdot\sum_{i,t}(\gamma_i' F_t+\varepsilon_{it}-c_i'F_t)^2\\
&= \frac{1}{NT}\cdot\sum_{i,t}\left[(\gamma_i'-c_i')F_t+\varepsilon_{it}\right]^2\\
&= \frac{1}{NT}\cdot\sum_{i,t}\left[(c_i'-\gamma_i')F_tF_t'(c_i-\gamma_i)-2(c_i'-\gamma_i')F_t\varepsilon_{it}+\varepsilon_{it}^2\right]\\
&= \frac{1}{N}\cdot\sum_{i=1}^{N}(c_i'-\gamma_i')\left(\frac{1}{T}\cdot\sum_{t=1}^{T}F_tF_t'\right)(c_i-\gamma_i)\\
&\quad -\frac{2}{N}\cdot\sum_{i=1}^{N}(c_i'-\gamma_i')\overline{(\varepsilon_i F)}+\frac{1}{NT}\cdot\sum_{i,t}\varepsilon_{it}^2
\end{aligned}$$

where $\overline{(\varepsilon_i F)}=\frac{1}{T}\cdot\sum_{t=1}^{T}\varepsilon_{it}F_t$. Since $\frac{1}{T}\cdot\sum_{t=1}^{T}F_tF_t'=\frac{1}{T}\cdot F'F=I_r$,

$$\widehat{Q}^2(c)=\frac{1}{N}\cdot\sum_{i=1}^{N}(c_i'-\gamma_i')(c_i-\gamma_i)-\frac{2}{N}\cdot\sum_{i=1}^{N}(c_i'-\gamma_i')\overline{(\varepsilon_i F)}+(const.)$$

By comparing this to $\widehat{Q}^2_{ref}(c)$ in (3.24), it is clear that the minimization problem (3.43) is a generalized version of the problem (3.25)[20]. In what follows, our goal is to prove that the estimator (3.43) provides the same result as the corresponding oracle estimator. This will be proven by Proposition 24, which is a generalized version of Lemma 8.

Regarding the above model, let us introduce some more notations. First,

$$\overline{c}_{G_j}=\frac{1}{N_j}\sum_{i=1}^{n}\mathbb{I}\left(i\in G_j\right)c_i$$

for any $\boldsymbol{c}=[c_{1,},\ldots,c_n]'$. Set $\boldsymbol{c}=\boldsymbol{m}+\boldsymbol{d}$ where the entries m_i of $\boldsymbol{m}$ are constant across each G_j and $\sum_{i=1}^{n}d_i\mathbb{I}\left(i\in G_j\right)=0$. Note that this decomposition is unique with, for i in G_j

$$\begin{aligned}
\boldsymbol{m}_i &= \overline{c}_{G_j}\\
\boldsymbol{d}_i &= c_i-\overline{c}_{G_j}.
\end{aligned}$$

Let $\boldsymbol{M}$ and $\boldsymbol{D}$ be the corresponding sets of $\boldsymbol{m}$'s and $\boldsymbol{d}$'s[21]. It is convenient to order the group index j according the $\overline{\gamma}_{G_j}$. Let (j) and $\gamma_{(j)}=\overline{\gamma}_{G_{(j)}}$ be

[20]The problem (3.43) considers the case of $dim(F_t)\geq 1$, whereas $dim(F_t)=1$ is assumed in the problem (3.25).

[21]By Assumption 18, $\boldsymbol{M}$ and $\boldsymbol{D}$ have finite size as well as Ω.

such

$$\gamma_{(1)} \leq \gamma_{(2)} \leq \cdots \leq \gamma_{(J_N)}$$

where J_N is the number of groups. Set also, recalling $\overline{\varepsilon}_i = \sum_{t=1}^{T} \varepsilon_{it}/T$,

$$\overline{\varepsilon}_{(j)} = \frac{1}{N_{(j)}} \sum_{i=1}^{n} \mathbb{I}\left(i \in G_{(j)}\right) \overline{\varepsilon}_i.$$

For $W = [w_{ik}, 1 \leq i, j \leq N]$ with $w_{ik} \geq 0$, set

$$\mathfrak{p}_W(\boldsymbol{c}) = \sum_{i=1}^{N} \sum_{k=1}^{N} w_{ik} \left|c_i - c_k\right|. \tag{Pen}$$

Let $\boldsymbol{W}$ be the set of W such that $w_{kj} \geq 0$ is constant when (i,k) belongs to $G_j \times G_l$ for each pair (G_j, G_l) with $j \neq l$. Set

$$W_{jl} = \sum_{i=1}^{N} \sum_{k=1}^{N} w_{ik} \mathbb{I}\left[(i,k) \in G_j \times G_l\right] = N_j N_l \overline{w}_{G_j \times G_l}$$

so that, setting $W_{(j)(l)} = W_{(jl)}$,

$$\mathfrak{p}_W(\boldsymbol{m}) = \sum_{1 \leq j \neq l \leq J_N} W_{(jl)} \left|m_{(j)} - m_{(l)}\right|.$$

For simplicity of the discussion, let us also impose the following unrestrictive assumption.

Assumption 18 *The estimator $\widehat{\gamma}$ exists in a space with finite size Ω $(\subset R^N)$.*

<u>Lemmas</u> In what follows, we introduce several lemmas: an inequality lemma, deviation lemmas, and two other ones related to the objective function.

Lemma 19 (Ineq) *Suppose $\underline{a} \geq -1$. Then for any $a \geq \underline{a}$ and $b \geq 0$*

$$\sqrt{1+a} + b \geq \sqrt{1 + a + 2b\sqrt{1+\underline{a}}}.$$

Proof of Lemma 19. As the two items of the inequality are nonnegative, the Lemma follows from

$$\left(\sqrt{1+a}+b\right)^2 = 1+2b\sqrt{1+a}+b^2 \geq 1+2b\sqrt{1+\underline{a}}. \qquad \square$$

Before introducing the deviation lemmas, let us define

$$\overline{\psi}_{il} \equiv \frac{1}{T} \cdot \sum_{t=1}^{T} F_{lt}\varepsilon_{it}$$

for $1 \leq l \leq r$. If $\{F_t\}_{t=1}^T$ is given, it holds that

- $E[\overline{\psi}_{il}|\{F_t\}] = 0$
- $Var[\overline{\psi}_{il}|\{F_t\}] = \dfrac{\sigma^2}{T}$
- $\overline{\psi}_{il}|\{F_t\} \sim N(0, \dfrac{\sigma^2}{T})$ (Note that I assume ε_{it} follows normal distribution.)
- $\overline{\psi}_{il}|\{F_t\}$ is *iid*.

for any given l. Then, Lemmas 20 and 21 are introduced (deviation lemmas).

Lemma 20 (Dev) *Suppose Assumptions 1(i), 2 and 5(i) hold. Then, for any given l and $\{F_t\}_{t=1}^T$,*

$$\lim_{N,T\to\infty} \mathbb{P}\left(\max_{1\leq i\leq N} \left|\overline{\psi}_{il}\right| \geq \sigma\sqrt{\frac{2\ln N}{T}} \,\middle|\, \{F_t\}\right) = 0,$$

$$\lim_{N,T\to\infty} \mathbb{P}\left(\max_{1\leq j\leq J_N} \sqrt{N_{(j)}T}\left|\overline{\psi}_{(j)l}\right| \geq \sigma\sqrt{2\ln J_N} \,\middle|\, \{F_t\}\right) = 0.$$

Proof of Lemma 20. Recall that, for $t > 0$,

$$\mathbb{P}\left(\left|\mathcal{N}\left(0,1\right)\right| \geq t\right) = \frac{2}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{u^2}{2}\right) du \leq \frac{2}{\sqrt{2\pi}} \int_t^\infty \frac{u}{t} \exp\left(-\frac{u^2}{2}\right) du$$

$$\leq \frac{2}{\sqrt{2\pi}} \left. -\frac{1}{t}\exp\left(-\frac{u^2}{2}\right)\right]_t^\infty = \frac{2}{\sqrt{2\pi}} \frac{\exp\left(-\frac{t^2}{2}\right)}{t}.$$

Since $\overline{\psi}_{il}|\{F_t\} \sim iidN(0, \frac{\sigma^2}{T})$, it follows for any T,

$$
\begin{aligned}
\mathbb{P}\left(\max_{1\le i\le N}|\overline{\psi}_{il}| \ge \sigma\sqrt{\frac{2\ln N}{T}}\,\middle|\,\{F_t\}\right) &\le \sum_{i=1}^{N}\mathbb{P}\left(|\overline{\psi}_{il}| \ge \sigma\sqrt{\frac{2\ln N}{T}}\,\middle|\,\{F_t\}\right)\\
&= N\mathbb{P}\left(\left|\mathcal{N}\left(0,\frac{\sigma^2}{T}\right)\right| \ge \sigma\sqrt{\frac{2\ln N}{T}}\right)\\
&\le N\mathbb{P}\left(|\mathcal{N}(0,1)| \ge \sigma\sqrt{2\ln N}\right)\\
&\le \frac{2N}{\sqrt{2\pi}}\frac{\exp\left(-\frac{\sqrt{2\ln N}^2}{2}\right)}{\sqrt{2\ln N}}\\
&\le \frac{2N}{\sqrt{2\pi}}\frac{1}{N\sqrt{2\ln N}}\\
&= \frac{1}{\sqrt{\pi\ln N}} \to 0. \qquad \square
\end{aligned}
$$

Lemma 21 (Dev2) *Suppose Assumptions 1(i), 2 and 5(i) hold. Then, for any given* l*,*

$$
\begin{aligned}
\lim_{N,T\to\infty}\mathbb{P}\left(\max_{1\le i\le N}|\overline{\psi}_{il}| \ge \sigma\sqrt{\frac{2\ln N}{T}}\right) &= 0,\\
\lim_{N,T\to\infty}\mathbb{P}\left(\max_{1\le j\le J_N}\sqrt{N_{(j)}T}\,|\overline{\psi}_{(j)l}| \ge \sigma\sqrt{2\ln J_N}\right) &= 0.
\end{aligned}
\tag{3.44}
$$

Proof (rough). The lemma can be proven by the following facts: Eq. (3.44) holds for any given $\{F_t\}$; ε_{it} is independent of $\{F_t\}$. $\square$

The next two lemmas are the ones related to the objective function.

Lemma 22 (Qdec) *Suppose Assumption 3 holds. Then, it holds for any* $\boldsymbol{c} = \boldsymbol{m} + \boldsymbol{d}$ *with* $(\boldsymbol{m}, \boldsymbol{d})$ *in* $\boldsymbol{M} \times \boldsymbol{D}$

$$
\begin{aligned}
&(i) & \widehat{Q}^2(\boldsymbol{c}) &= \widehat{Q}^2(\boldsymbol{m}) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((d_i'F_t)^2 - 2(d_i'F_t)\cdot\varepsilon_{it}\right),\\
&(ii) & \frac{1}{T}\cdot\left|\sum_{i=1}^{N}\sum_{t=1}^{T}(d_i'F_t)\cdot\varepsilon_{it}\right| &\le \sum_{j=1}^{J_N}\frac{1}{N_j}\cdot\sum_{(i,k)\in G_j\times G_j}\sum_{l=1}^{r}|d_{i(l)} - d_{k(l)}|\cdot\max_{i,l}\left|\overline{(\varepsilon F)}_{il}\right|.
\end{aligned}
$$

Decompose $\widehat{\boldsymbol{\gamma}}$ *into* $\widehat{\boldsymbol{m}} + \widehat{\boldsymbol{d}}$*. Then, under Assumptions 2 and 18,*

$$
(iii) \quad \widehat{Q}^2(\widehat{\boldsymbol{m}}) \ge \sigma^2 + o(1)
$$

$$(iv) \qquad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{(\widehat{d}_i'F_t)^2 - 2(\widehat{d}_i'F_t)\cdot\varepsilon_{it}\right\} \geq -\frac{r\widehat{Q}^2\left(\widehat{\boldsymbol{m}}\right)}{T}\left(1+o(1)\right)$$

with a probability tending to 1.

Proof of Lemma 22.

(i) Indeed,

$$\widehat{Q}^2\left(\boldsymbol{c}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\gamma_i'F_t - m_i'F_t - d_i'F_t\right)^2 + 2\left(\gamma_i'F_t - m_i'F_t - d_i'F_t\right)\cdot\varepsilon_{it} + \varepsilon_{it}^2\right\}$$

with

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\gamma_i'F_t - m_i'F_t - d_i'F_t\right)^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\gamma_i'F_t - m_i'F_t\right)^2 + (d_i'F_t)^2\right\}\\ &\quad -\frac{2}{NT}\sum_{j=1}^{J_N}\sum_{t=1}^{T}\left(\overline{\gamma}_{G_j}'F_t - \overline{m}_{G_j}'F_t\right)\cdot\sum_{i\in G_j}d_i'F_t\\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\gamma_i'F_t - m_i'F_t\right)^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(d_i'F_t)^2.\end{aligned}$$

Hence

$$\begin{aligned}\widehat{Q}^2\left(\boldsymbol{c}\right) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\gamma_i'F_t - m_i'F_t\right)^2 + 2\left(\gamma_i'F_t - m_i'F_t\right)\varepsilon_{it} + \varepsilon_{it}^2\right\}\\ &\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(d_i'F_t)^2 - 2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(d_i'F_t)\cdot\varepsilon_{it}\\ &= \widehat{Q}^2\left(\boldsymbol{m}\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(d_i'F_t)^2 - 2(d_i'F_t)\cdot\varepsilon_{it}\right].\end{aligned}$$

(ii) For the second equality, it holds since $\overline{d}_{G_j} = \frac{1}{N_j}\sum_{k\in G_j} d_k = 0$

$$\begin{aligned}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}(d_i'F_t)\cdot\varepsilon_{it} &= \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{r}d_{i(l)}F_{lt}\varepsilon_{it} = \sum_{i=1}^{N}\sum_{l=1}^{r}d_{i(l)}\cdot\overline{(\varepsilon F)_{il}}\\ &= \sum_{j=1}^{J_N}\sum_{i\in G_j}\sum_{l=1}^{r}\left(d_{i(l)} - \overline{d}_{G_j(l)}\right)\cdot\overline{(\varepsilon F)_{il}}\\ &= \sum_{j=1}^{J_N}\frac{1}{N_j}\sum_{(i,k)\in G_j}\sum_{l=1}^{r}\left(d_{i(l)} - d_{k(l)}\right)\cdot\overline{(\varepsilon F)_{il}}\\ &\leq \sum_{j=1}^{J_N}\frac{1}{N_j}\sum_{(i,k)\in G_j}\sum_{l=1}^{r}\left(d_{i(l)} - d_{k(l)}\right)\cdot\left|\overline{(\varepsilon F)_{il}}\right|\\ &\leq \sum_{j=1}^{J_N}\frac{2}{N_j}\sum_{(i,k)\in G_j}\sum_{l=1}^{r}\left|d_{i(l)} - d_{k(l)}\right|\cdot\max_{i,l}\left|\overline{(\varepsilon F)_{il}}\right|\end{aligned}$$

(iii)

$$\begin{aligned}
\widehat{Q}^2(\widehat{\gamma}) &= \frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\widehat{\gamma}_i'F_t)^2 \\
&= \frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{it}+(\gamma_i'-\widehat{\gamma}_i')F_t\right]^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2+\frac{2}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(\gamma_i'-\widehat{\gamma}_i')F_t\varepsilon_{it}+\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}[(\gamma_i'-\widehat{\gamma}_i')F_t]^2 \\
&\geq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2+\frac{2}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(\gamma_i'-\widehat{\gamma}_i')F_t\varepsilon_{it} \\
&= \sigma^2+O_p((NT)^{-1/2})+\frac{2}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(\gamma_i'-\widehat{\gamma}_i')F_t\varepsilon_{it}
\end{aligned}$$

Here,

$$\begin{aligned}
&\left|\frac{2}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(\gamma_i'-\widehat{\gamma}_i')F_t\varepsilon_{it}\right| \\
&\quad= \left|\frac{2}{N}\cdot\sum_{i=1}^{N}(\gamma_i'-\widehat{\gamma}_i')\left[\frac{1}{T}\cdot\sum_{t=1}^{T}F_t\varepsilon_{it}\right]\right| \\
&\quad= \left|\frac{2}{N}\cdot\sum_{l=1}^{r}\sum_{i=1}^{N}(\gamma_{i(l)}-\widehat{\gamma}_{i(l)})\left[\frac{1}{T}\cdot\sum_{t=1}^{T}F_{lt}\varepsilon_{it}\right]\right| \\
&\quad\leq \frac{2}{N}\cdot\sum_{l=1}^{r}\sum_{i=1}^{N}|\gamma_{i(l)}-\widehat{\gamma}_{i(l)}|\cdot\left|\frac{1}{T}\cdot\sum_{t=1}^{T}F_{lt}\varepsilon_{it}\right| \\
&\quad\leq \frac{2}{N}\cdot\sum_{l=1}^{r}\sum_{i=1}^{N}|\gamma_{i(l)}-\widehat{\gamma}_{i(l)}|\cdot\max_{1\leq i\leq N}|\overline{\psi}_{il}| \\
&\quad\leq 4\sigma\cdot\sqrt{\frac{2\ln N}{T}}\cdot\left(\frac{1}{N}\cdot\sum_{l=1}^{r}\sum_{i=1}^{N}|\gamma_{i(l)}-\widehat{\gamma}_{i(l)}|\right) \qquad \text{(from Lemma 21)}
\end{aligned} \tag{3.45}$$

with a probability tending to 1. From Assumptions 5(i) and 18, this suggests

$$\left|\frac{2}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}(\gamma_i'-\widehat{\gamma}_i')F_t\varepsilon_{it}\right|\leq o(1)$$

Thus,

$$\widehat{Q}^2(\widehat{\gamma})\geq\sigma^2+o(1)\implies\widehat{Q}^2(\widehat{m})\geq\sigma^2+o(1)$$

with a probability tending to 1. $\square$

(iv) First,

$$\begin{aligned}
&\frac{1}{T}\cdot\sum_{t=1}^{T}(\widehat{d}_i'F_t)^2=\frac{1}{T}\cdot\sum_{t=1}^{T}\widehat{d}_i'F_tF_t'\widehat{d}_i=\widehat{d}_i'\left(\frac{1}{T}\cdot\sum_{t=1}^{T}F_tF_t'\right)\widehat{d}_i=\widehat{d}_i'\widehat{d}_i=\sum_{l=1}^{r}\widehat{d}_{i(l)}^2 \\
&\frac{1}{T}\cdot\sum_{t=1}^{T}(\widehat{d}_i'F_t)\varepsilon_{it}=\frac{1}{T}\cdot\sum_{l=1}^{r}\sum_{t=1}^{T}\widehat{d}_{i(l)}F_{lt}\varepsilon_{it}=\sum_{l=1}^{r}\widehat{d}_{i(l)}\left(\frac{1}{T}\cdot\sum_{t=1}^{T}F_{lt}\varepsilon_{it}\right)=\sum_{l=1}^{r}\widehat{d}_{i(l)}\overline{\psi}_{il}
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{(\widehat{d}_i'F_t)^2-2\cdot(\widehat{d}_i'F_t)\varepsilon_{it}\right\} &= \frac{1}{N}\cdot\sum_{i=1}^{N}\sum_{l=1}^{r}\left(\widehat{d}_{i(l)}^2-2\widehat{d}_{i(l)}\overline{\psi}_{il}\right)\\
&\geq -\frac{1}{N}\cdot\sum_{i=1}^{N}\sum_{l=1}^{r}\left(\overline{\psi}_{il}\right)^2\\
&= -\sum_{l=1}^{r}\left[\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\overline{\psi}_{il}\right)^2\right]
\end{aligned}
\tag{3.46}
$$

Here, it is easy to see

$$
E[\left(\overline{\psi}_{il}\right)^2|\{F_t\},l]=\frac{\sigma^2}{T}
$$

for any of given $\{F_t\}$ and l. Due to this,

$$
\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\overline{\psi}_{il}\right)^2 \quad \xrightarrow{P} \quad \frac{\sigma^2}{T} \tag{3.47}
$$

which is obtained by using the fact that for any of given $\{F_t\}$ and l, $(\overline{\psi}_{il})^2$ in the summation in the LHS can be regarded as i.i.d. series. Using (3.46), (3.47) and the sublemma (iii),

$$
\begin{aligned}
\frac{1}{NT}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{(\widehat{d}_i'F_t)^2-2\cdot(\widehat{d}_i'F_t)\varepsilon_{it}\right\} &\geq -\frac{r\sigma^2}{T}\\
&\geq -\frac{r\widehat{Q}^2(\widehat{m})}{T}\cdot(1+o(1))
\end{aligned}
$$

with a probability tending to 1. $\square$

Lemma 23 (Pdec) *Let $\mathfrak{p}_W(\cdot)$ be as in (Pen). It holds for all W in $\boldsymbol{W}$ and all $(\boldsymbol{m},\boldsymbol{d})$ in $\boldsymbol{M}\times\boldsymbol{D}$*

$$
\mathfrak{p}_W(\boldsymbol{m}+\boldsymbol{d})\geq\mathfrak{p}_W(\boldsymbol{m})+\sum_j\sum_{(i,k)\in G_j\times G_j}w_{ik}\left|d_i-d_k\right|.
$$

Proof of Lemma 23. Since m_i is constant when i belongs to G_j it holds

$$\mathfrak{p}_W(\boldsymbol{m}+\boldsymbol{d}) = \sum_j \sum_{(i,k)\in G_j\times G_j} w_{ik}\,|d_i - d_k| + \sum_{j\neq l}\sum_{(i,k)\in G_j\times G_l} w_{ik}\,|m_i - m_k - (d_i - d_k)|\,.$$

Since $m_i - m_k$ is constant across $G_j \times G_l$, the triangular inequality, $\sum_{i\in G_j} d_i = \sum_{k\in G_l} d_k = 0$ and the definition of $\boldsymbol{W}$ with $w_{ik} = \overline{w}_{G_j\times G_l}$ for all (i,k) in $G_j \times G_l$ give

$$\begin{aligned}\sum_{(i,k)\in G_j\times G_l} w_{ik}\,|m_i + d_i - (m_k + d_k)| &= \overline{w}_{G_j\times G_l} \sum_{(i,k)\in G_j\times G_l} |m_i - m_k - (d_i - d_k)| \\ &\geq \overline{w}_{G_j\times G_l}\left|\sum_{(i,k)\in G_j\times G_l} (m_i - m_k) - 0\right| \\ &= \sum_{(i,k)\in G_j\times G_l} w_{ik}\,|m_i - m_k|\,. \qquad \square\end{aligned}$$

Classification proposition for the Lasso estimator In this paragraph, we will prove the classification proposition for the Lasso estimator (Proposition 24), which is a generalized version of Lemma 8. Let $\widehat{W} = \widehat{W}_{NT}$ be a stochastic sequence in $\boldsymbol{W}$ and define, for $\widehat{\lambda} = \lambda_{NT} \geq 0$

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}\left(\widehat{W},\widehat{\lambda}\right) &= \arg\min_{\boldsymbol{c}}\left\{\widehat{Q}(\boldsymbol{c}) + \widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}(\boldsymbol{c})\right\}, \\ \widehat{\boldsymbol{\mu}}\left(\widehat{W},\widehat{\lambda}\right) &= \arg\min_{\boldsymbol{m}\in\boldsymbol{M}}\left\{\widehat{Q}(\boldsymbol{m}) + \widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}(\boldsymbol{m})\right\}.\end{aligned}$$

Proposition 24 (Classification) *Suppose Assumptions 1(i), 2, 3, 5(i) and 18 hold. Assume there is a $\delta > 0$ such that*

$$\widehat{\lambda}\min_j \min_{(i,k)\in G_j\times G_j}\{N_j\widehat{w}_{ik}\} \geq \frac{1+\delta+o_{\mathbb{P}}(1)}{N}\sqrt{\frac{2\ln N}{T}}.$$

Then $\widehat{\boldsymbol{\gamma}}\left(\widehat{W},\widehat{\lambda}\right) = \widehat{\boldsymbol{\mu}}\left(\widehat{W},\widehat{\lambda}\right)$ with a probability tending to 1.

Proof of Proposition 24. Set $\widehat{\boldsymbol{\gamma}}\left(\widehat{W},\widehat{\lambda}\right) = \widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{\mu}}\left(\widehat{W},\widehat{\lambda}\right) = \widehat{\boldsymbol{\mu}}$ for the sake of brevity and decompose $\widehat{\boldsymbol{\gamma}}$ into $\widehat{\boldsymbol{m}}+\widehat{\boldsymbol{d}}$. Then the definition of these estimators, Lemmas 22(i), and 23 give

$$
\begin{aligned}
&\widehat{Q}\left(\widehat{\boldsymbol{\mu}}\right)+\widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}\left(\widehat{\boldsymbol{\mu}}\right)\geq\widehat{Q}\left(\widehat{\boldsymbol{\gamma}}\right)+\widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}\left(\widehat{\boldsymbol{\gamma}}\right)\\
&\quad\geq\sqrt{\widehat{Q}^2\left(\widehat{\boldsymbol{m}}\right)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\widehat{d}_i'F_t)^2-2(\widehat{d}_i'F_t)\cdot\varepsilon_{it}\right)}\\
&\qquad\qquad+\widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}\left(\widehat{\boldsymbol{m}}\right)+\ \widehat{\lambda}\cdot\sum_j\sum_{(i,k)\in G_j\times G_j}\widehat{w}_{ik}\cdot\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|\\
&\quad=\widehat{Q}\left(\widehat{\boldsymbol{m}}\right)+\widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}\left(\widehat{\boldsymbol{m}}\right)+\widehat{Q}\left(\widehat{\boldsymbol{m}}\right)\left\{\sqrt{1+A\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)}+B\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)-1\right\}
\end{aligned}
$$

Note that

$$
\begin{aligned}
A\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)&=\frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\widehat{d}_i'F_t)^2-2(\widehat{d}_i'F_t)\cdot\varepsilon_{it}\right)}{\widehat{Q}^2\left(\widehat{\boldsymbol{m}}\right)}\geq-\frac{r(1+o(1))}{T},\\
B\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)&=\widehat{\lambda}\cdot\frac{\sum_j\sum_{(i,k)\in G_j\times G_j}\widehat{w}_{ik}\cdot\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|}{\widehat{Q}\left(\widehat{\boldsymbol{m}}\right)}.
\end{aligned}
$$

where we used Lemma 22(iv) in the first line. As $\widehat{Q}\left(\widehat{\boldsymbol{\mu}}\right)+\widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}\left(\widehat{\boldsymbol{\mu}}\right)\leq\widehat{Q}\left(\widehat{\boldsymbol{m}}\right)+\widehat{\lambda}\cdot\mathfrak{p}_{\widehat{W}}\left(\widehat{\boldsymbol{m}}\right)$, Lemmas 19 and 22(iv) give, with a probability tending to 1,

$$
\begin{aligned}
0&\geq\sqrt{1+A\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)}+B\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)-1\\
&\geq\sqrt{1+A\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)+2B\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)\sqrt{1-\frac{r(1+o(1))}{T}}}-1
\end{aligned}
$$

and then

$$
0\geq A\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)+2B\left(\widehat{\boldsymbol{m}},\widehat{\boldsymbol{d}}\right)\sqrt{1-\frac{r(1+o(1))}{T}}.
$$

It then follows, by Lemma 22(iii)

$$
\begin{aligned}
0&\geq\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\widehat{d}_i'F_t)^2-2(\widehat{d}_i'F_t)\cdot\varepsilon_{it}\right)\\
&\qquad+2\widehat{\lambda}\cdot\widehat{Q}\left(\widehat{\boldsymbol{m}}\right)\cdot\sqrt{1-o_{\mathbb{P}}\left(1\right)}\cdot\sum_j\sum_{(i,k)\in G_j\times G_j}\widehat{w}_{ik}\cdot\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|
\end{aligned}
$$

$$
\begin{aligned}
&\geq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\widehat{d}_i' F_t)^2 - 2(\widehat{d}_i' F_t)\cdot\varepsilon_{it}\right)\\
&\qquad + 2\widehat{\lambda}\cdot\sigma\cdot(1-o_{\mathbb{P}}(1))\cdot\sum_j\sum_{(i,k)\in G_j\times G_j}\widehat{w}_{ik}\cdot\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|.
\end{aligned}
$$

Bounding $\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{d}_i' F_t)\cdot\varepsilon_{it}$ from below as in Lemma 22(ii) gives, with an $o_{\mathbb{P}}(1)$ remainder term independent of i, j and k,

$$
\begin{aligned}
0 \quad \geq \quad & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{d}_i' F_t)^2\\
&+\frac{2}{N}\sum_j\frac{1}{N_j}\sum_{(i,k)\in G_j\times G_j}\left\{(1-o_{\mathbb{P}}(1))\,\widehat{\lambda}NN_j\sigma\widehat{w}_{ik}-\max_{(i,l)}\left|\overline{(\varepsilon F)}_{il}\right|\right\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\times\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|.
\end{aligned}
$$

This gives by Lemma 21

$$
\begin{aligned}
0 \geq\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{d}_i' F_t)^2\\
&\qquad + \frac{2\sigma}{N}\left((1-o_{\mathbb{P}}(1))\,\widehat{\lambda}N\min_j\min_{(i,k)\in G_j\times G_j}\{N_j\widehat{w}_{ik}\}-\sqrt{\frac{2\ln N}{T}}\right)\\
&\qquad\qquad\qquad\times\sum_j\frac{1}{N_j}\sum_{(i,k)\in G_j\times G_j}\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|
\end{aligned}
$$

and, under the conditions of the Proposition

$$
\begin{aligned}
0 \geq\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{d}_i' F_t)^2\\
&\qquad + (\delta+o_{\mathbb{P}}(1))\,\frac{2\sigma}{N}\sqrt{\frac{2\ln N}{T}}\sum_j\frac{1}{N_j}\sum_{(i,k)\in G_j\times G_j}\sum_{l=1}^{r}\left|\widehat{d}_{i(l)}-\widehat{d}_{k(l)}\right|.
\end{aligned}
$$

As $\delta > 0$ and $|F_l| \neq 0$ for any l, this implies that $\widehat{d}_{i(l)}^2$ must be equal to 0 for any (i,l), with a probability tending to 1. Hence $\mathbb{P}(\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{m}}) = 1 - o(1)$ and then $\mathbb{P}(\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\mu}}) = 1 - o(1)$ by definition of $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{m}}$ and $\widehat{\boldsymbol{\mu}}$. □

3.8 Appendices

3.8.1 MC sample generation

Generation of latent factors Denote the latent factors by

$$F^{(j)} \equiv [F_{j1}, F_{j2}, \ ... \ , F_{jT}]' \quad (j = 1, 2, 3)$$

In the Monte Carlo experiments, these vectors are generated by the following procedure.

- First, generate H_{jt} $(j = 1, 2, 3)$ by

$$\begin{aligned}
&H_{jt} = 0.5 \cdot H_{j,t-1} + \xi_{jt} \quad (t = -49, ..., 1, ..., T) \\
&H_{j,-50} = 0 \\
&\xi_{jt} \sim iidN(0, \sigma_\xi^2) \text{ with } \sigma_\xi^2 = 1 - 0.5^2
\end{aligned}$$

  which obtains $H^{(j)} = [H_{j1}, ..., H_{jT}]'$.
- Define $F^{(1)}$ by

$$F^{(1)} = \frac{H^{(1)}}{\sqrt{H^{(1)\prime} H^{(1)}/T}}$$

- Define $F^{(2)}$ by

$$\begin{aligned}
\widetilde{H}^{(2)} &\equiv H^{(2)} - \frac{F^{(1)\prime} H^{(2)}}{F^{(1)\prime} F^{(1)}} \cdot F^{(1)} \\
F^{(2)} &= \frac{\widetilde{H}^{(2)}}{\sqrt{\widetilde{H}^{(2)\prime} \widetilde{H}^{(2)}/T}}
\end{aligned}$$

- Define $F^{(3)}$ by

$$\begin{aligned}
\widetilde{H}^{(3)} &\equiv H^{(3)} - \frac{F^{(1)\prime} H^{(3)}}{F^{(1)\prime} F^{(1)}} \cdot F^{(1)} - \frac{F^{(2)\prime} H^{(3)}}{F^{(2)\prime} F^{(2)}} \cdot F^{(2)} \\
F^{(3)} &= \frac{\widetilde{H}^{(3)}}{\sqrt{\widetilde{H}^{(3)\prime} \widetilde{H}^{(3)}/T}}
\end{aligned}$$

The loadings γ_i In the baseline setup, the loadings γ_i are defined by

$$\gamma_i = \gamma_i^{(base)} \equiv \begin{cases} (2.0, -0.5)' & (i \in G_1) \\ (1.5, 0.0)' & (i \in G_2) \\ (1.0, 1.0)' & (i \in G_3) \end{cases}$$

It should be emphasized that this satisfies the identification restriction, i.e.

$$\frac{1}{N} \cdot \sum_{i=1}^{N} \gamma_i \gamma_i' \text{ is diagonal.} \tag{3.48}$$

The loadings Γ_i In any cases of the MC experiments, the loadings Γ_i have the following form:

$$\Gamma_i^{(1)} = \begin{pmatrix} iidN(\mu_\Gamma, \sigma_\Gamma^2) \\ 0 \\ iidN(0, \sigma_\Gamma^2) \end{pmatrix} \quad , \quad \Gamma_i^{(2)} = \begin{pmatrix} iidN(0, \sigma_\Gamma^2) \\ 0 \\ iidN(\mu_\Gamma, \sigma_\Gamma^2) \end{pmatrix}$$

With this form, it is easy to show that

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \cdot \left(\sum_{i=1}^{N} \Gamma_i^{(1)} \Gamma_i^{(1)'} \right) &= \begin{pmatrix} \sigma_\Gamma^2 - \mu_\Gamma^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_\Gamma^2 \end{pmatrix} \quad \text{(diagonal)} \\ \lim_{N\to\infty} \frac{1}{N} \cdot \left(\sum_{i=1}^{N} \Gamma_i^{(2)} \Gamma_i^{(2)'} \right) &= \begin{pmatrix} \sigma_\Gamma^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_\Gamma^2 - \mu_\Gamma^2 \end{pmatrix} \quad \text{(diagonal)} \end{aligned} \tag{3.49}$$

3.8.2 Optimization of penalty parameter λ

The penalty parameter λ is optimized by using BIC (Bayesian information criterion). BIC is defined by

$$BIC = log(\widehat{L}_\lambda) + \left|\widehat{S}_\lambda\right| \cdot \frac{log(NT)}{NT} \tag{3.50}$$

Note that

$$\widehat{L}_\lambda = \frac{1}{NT} \cdot \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \widehat{\beta}' X_{it} - \widehat{\alpha}_i' \overline{Z}_t)^2$$

where $\widehat{\alpha_i}$ and $\widehat{\beta}$ are the Lasso-based estimator. $\left|\widehat{S}_\lambda\right|$ is the identified number of parameters in the model, whose definition is

$$\begin{aligned} &\widehat{S}_\lambda = \{\widehat{\beta}_1, ..., \widehat{\beta}_k, \widehat{S}_{\alpha,1}, ..., \widehat{S}_{\alpha,k+1}\} \\ &\widehat{S}_{\alpha,l} = \{i : 1 \le i \le N, |\widehat{\alpha}_{il} - \widehat{\alpha}_{jl}| \neq 0 \text{ for all } j \text{ with } j < i\} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad (l = 1, ..., k+1) \end{aligned}$$

3.8.3 Other proofs

Derivation of Eq. (3.35)

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\Delta_{it}\widetilde{y}_{it} &= \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t\widetilde{y}_{it} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t y_{it} \\
&\quad - \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t \\
&\qquad \left[\frac{1}{N_j}\cdot\overline{Z}_t'\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{s=1}^{T}\sum_{k\in G_j}\overline{Z}_s y_{ks}\right)\right] \\
&= \sum_{j=1}^{J_N}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i\in G_j}\overline{Z}_t y_{it}\right) \\
&\quad - \sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j^2}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{t=1}^{T}\overline{Z}_t\overline{Z}_t'\right) \\
&\qquad \left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{s=1}^{T}\sum_{k\in G_j}\overline{Z}_s y_{ks}\right) \\
&= \sum_{j=1}^{J_N}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i\in G_j}\overline{Z}_t y_{it}\right) \\
&\quad - \sum_{j=1}^{J_N}\frac{N_j}{N_j^2}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{s=1}^{T}\sum_{k\in G_j}\overline{Z}_s y_{ks}\right) \\
&= \sum_{j=1}^{J_N}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{t=1}^{T}\sum_{i\in G_j}\overline{Z}_t y_{it}\right) \\
&\quad - \sum_{j=1}^{J_N}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{s=1}^{T}\sum_{k\in G_j}\overline{Z}_s y_{ks}\right) \\
&= 0
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\Delta_{it}\widetilde{X}_{it}' &= \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t X_{it}' \\
&\quad - \sum_{t=1}^{T}\sum_{j=1}^{J_N}\sum_{i\in G_j}\frac{1}{N_j}\cdot\left(\sum_{s=1}^{T}\sum_{k\in G_j}X_{ks}\overline{Z}_s'\right)\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\overline{Z}_t \\
&\qquad \left[\frac{1}{N_j}\cdot\overline{Z}_t'\left(\sum_{s=1}^{T}\overline{Z}_s\overline{Z}_s'\right)^{-1}\left(\sum_{s=1}^{T}\sum_{k\in G_j}\overline{Z}_s X_{ks}'\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=1}^{J_N} \frac{1}{N_j} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{t=1}^{T} \sum_{i \in G_j} \overline{Z}_t X_{it}' \right) \\
&\quad - \sum_{j=1}^{J_N} \sum_{i \in G_j} \frac{1}{N_j^2} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{t=1}^{T} \overline{Z}_t \overline{Z}_t' \right) \\
&\qquad \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{s=1}^{T} \sum_{k \in G_j} \overline{Z}_s X_{ks}' \right) \\
&= \sum_{j=1}^{J_N} \frac{1}{N_j} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{t=1}^{T} \sum_{i \in G_j} \overline{Z}_t X_{it}' \right) \\
&\quad - \sum_{j=1}^{J_N} \frac{N_j}{N_j^2} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{s=1}^{T} \sum_{k \in G_j} \overline{Z}_s X_{ks}' \right) \\
&= \sum_{j=1}^{J_N} \frac{1}{N_j} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{t=1}^{T} \sum_{i \in G_j} \overline{Z}_t X_{it}' \right) \\
&\quad - \sum_{j=1}^{J_N} \frac{1}{N_j} \cdot \left(\sum_{s=1}^{T} \sum_{k \in G_j} X_{ks} \overline{Z}_s' \right) \left(\sum_{s=1}^{T} \overline{Z}_s \overline{Z}_s' \right)^{-1} \left(\sum_{s=1}^{T} \sum_{k \in G_j} \overline{Z}_s X_{ks}' \right) \\
&= 0 \quad \square
\end{aligned}
$$

3.8.4 Order estimation

$\frac{1}{T} \cdot \left(\sum_{t=1}^{T} \overline{U}_t \overline{U}_t' \right)$ etc.

By CLT,

$$
\frac{1}{T} \cdot \sum_{t=1}^{T} \overline{V}_t \overline{\varepsilon}_t \sim N\left(0,\ \frac{\sigma_\varepsilon^2 \sigma_v^2}{N^2 T} \cdot I_k\right)
$$

Therefore,

$$
\frac{1}{T} \cdot \sum_{t=1}^{T} \overline{V}_t \overline{\varepsilon}_t = O\left(\frac{1}{\sqrt{N^2 T}}\right) \tag{3.51}
$$

As for $\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{V}_t\overline{V}_t'$ and $\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{\varepsilon}_t\overline{\varepsilon}_t$,

$$\begin{aligned}
\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{V}_t\overline{V}_t' &= \frac{1}{N^2T}\cdot\sum_{i,k,t}V_{it}V_{kt}' \\
&= \left[\frac{1}{N^2T}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}V_{it}V_{it}'\right]+\left[\frac{2}{N^2T}\cdot\sum_{i=1}^{N}\sum_{k>i}\sum_{t=1}^{T}V_{it}V_{kt}'\right] \\
&= \frac{1}{N}\cdot\left[\sigma_v^2 I_k+O\left(\frac{1}{\sqrt{NT}}\right)\right]+O\left(\frac{1}{\sqrt{N^2T}}\right) \qquad \text{(by CLT)} \\
&= O\left(\frac{1}{N}\right)
\end{aligned} \tag{3.52}$$

$$\begin{aligned}
\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{\varepsilon}_t\overline{\varepsilon}_t &= \frac{1}{N^2T}\cdot\sum_{i,k,t}\varepsilon_{it}\varepsilon_{kt} \\
&= \left(\frac{1}{N^2T}\cdot\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2\right)+\left(\frac{2}{N^2T}\cdot\sum_{i=1}^{N}\sum_{k>i}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\right) \\
&= \frac{1}{N}\cdot\left(\sigma^2+O\left(\frac{1}{\sqrt{NT}}\right)\right)+O\left(\frac{1}{\sqrt{N^2T}}\right) \qquad \text{(by CLT)} \\
&= O\left(\frac{1}{N}\right)
\end{aligned} \tag{3.53}$$

Using (3.51)-(3.53), it can be obtained

$$\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{\varepsilon}_t\overline{U}_t' = O\left(\frac{1}{N}\right) \tag{3.54}$$

$$\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{V}_t\overline{U}_t' = O\left(\frac{1}{N}\right) \tag{3.55}$$

$$\frac{1}{T}\cdot\sum_{t=1}^{T}\overline{U}_t\overline{U}_t' = O\left(\frac{1}{N}\right) \tag{3.56}$$

because $\overline{U}_t$ is a linear function of $\overline{\varepsilon}_t$ and $\overline{V}_t$.

$\frac{1}{T} \cdot \left(\sum_{t=1}^{T} F_t \overline{U}'_t \right)$ **etc.**

Using the fact that F_t is independent of (ε_{it}, V_t), it is easy to show

$$E\left[\frac{1}{NT} \cdot \sum_{i,t} F_t \varepsilon_{it} \right] = 0 \ , \quad Var\left[\frac{1}{NT} \cdot \sum_{i,t} F_t \varepsilon_{it} \right] = \frac{\sigma_\varepsilon^2}{NT} \cdot I_m$$
$$E\left[\frac{1}{NT} \cdot \sum_{i,t} F_t V'_{it} \right] = 0 \ , \quad Var\left[\frac{1}{NT} \cdot \sum_{i,t} F_t V'_{it} \right] = \frac{k\sigma_v^2}{NT} \cdot I_m$$

From this, let us suppose

$$\frac{1}{T} \cdot \sum_{t=1}^{T} F_t \overline{\varepsilon}_t = \frac{1}{NT} \cdot \sum_{i,t} F_t \varepsilon_{it} = O\left(\frac{1}{\sqrt{NT}} \right) \tag{3.57}$$

$$\frac{1}{T} \cdot \sum_{t=1}^{T} F_t \overline{V}'_t = \frac{1}{NT} \cdot \sum_{i,t} F_t V_{it} = O\left(\frac{1}{\sqrt{NT}} \right) \tag{3.58}$$

which yield

$$\frac{1}{T} \cdot \sum_{t=1}^{T} F_t \overline{U}'_t = O\left(\frac{1}{\sqrt{NT}} \right) \tag{3.59}$$

$\frac{1}{T} \cdot \left(\sum_{t=1}^{T} \varepsilon_{it} \overline{U}'_t \right)$ **,** $\frac{1}{T} \cdot \left(\sum_{t=1}^{T} V_{it} \overline{U}'_t \right)$ **etc.**

First,

$$\begin{aligned} \frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{it} \overline{\varepsilon}_t &= \frac{1}{NT} \cdot \sum_{k,t} \varepsilon_{it} \varepsilon_{kt} \\ &= \frac{1}{NT} \cdot \sum_{t=1}^{T} \varepsilon_{it}^2 + \frac{2}{NT} \cdot \sum_{t=1}^{T} \sum_{k>i} \varepsilon_{it} \varepsilon_{kt} \\ &= \frac{1}{N} \cdot \left(\sigma_\varepsilon^2 + O\left(\frac{1}{\sqrt{T}} \right) \right) + O\left(\frac{1}{\sqrt{NT}} \right) \quad \text{(by CLT)} \\ &= O\left(\frac{1}{N} \right) + O\left(\frac{1}{\sqrt{NT}} \right) \end{aligned} \tag{3.60}$$

$$\frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{it} \overline{V}'_t = \frac{1}{NT} \cdot \sum_{k,t} \varepsilon_{it} V'_{kt} = O\left(\frac{1}{\sqrt{NT}} \right) \tag{3.61}$$

which yields

$$\frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{it} \overline{U}'_t = O\left(\frac{1}{N} \right) + O\left(\frac{1}{\sqrt{NT}} \right) \tag{3.62}$$

Also,

$$\frac{1}{T}\cdot\sum_{t=1}^{T}V_{it}\overline{\varepsilon}_t=\frac{1}{NT}\cdot\sum_{k,t}V_{it}\varepsilon_{kt}=O\left(\frac{1}{\sqrt{NT}}\right) \tag{3.63}$$

$$\begin{aligned}\frac{1}{T}\cdot\sum_{t=1}^{T}V_{it}\overline{V}'_t &= \frac{1}{NT}\cdot\sum_{k,t}V_{it}V'_{kt}\\ &= \frac{1}{NT}\cdot\sum_{t=1}^{T}V_{it}V'_{it}+\frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{k\neq i}V_{it}V'_{kt}\\ &= \frac{1}{N}\cdot\left[\sigma^2 I+O\left(\frac{1}{\sqrt{T}}\right)\right]+O\left(\frac{1}{\sqrt{NT}}\right)\\ &= O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{NT}}\right)\end{aligned} \tag{3.64}$$

Therefore,

$$\frac{1}{T}\cdot\sum_{t=1}^{T}V_{it}\overline{U}'_t=O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{NT}}\right) \tag{3.65}$$

$\psi_1^{(a)}$ (in Lemma 10)

$$\begin{aligned}\psi_1^{(a)} &= \frac{1}{T}\cdot\sum_{t=1}^{T}(\alpha'_iC'+\beta'\Gamma'_i)F_tU'_t\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\alpha'_iU_tF'_tC\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\alpha'_iU_tU'_t\\ &\quad+\frac{1}{T}\cdot\sum_{t=1}^{T}\beta'V_{it}F'_tC\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\beta'V_{it}U'_t\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\eta_{it}U'_t\\ &= \frac{1}{T}\cdot\sum_{t=1}^{T}(\alpha'_iC'+\beta'\Gamma'_i)F_tU'_t\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\alpha'_iU_tF'_tC\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\alpha'_iU_tU'_t\\ &\quad+\frac{1}{T}\cdot\sum_{t=1}^{T}\beta'V_{it}F'_tC\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\beta'V_{it}U'_t\ +\ \frac{1}{T}\cdot\sum_{t=1}^{T}\varepsilon_{it}U'_t\\ &\quad-\frac{1}{T}\cdot\sum_{t=1}^{T}\alpha'_iU_tU'_t\end{aligned} \tag{3.66}$$

As for the 1st and 2nd terms of RHS,

$$\begin{aligned}\text{(1st-term)} &= (\alpha'_iC'+\beta'\Gamma'_i)\left[\frac{1}{T}\cdot\sum_{t=1}F_tU'_t\right]\\ \text{(2nd-term)} &= \alpha'_i\left(\frac{1}{T}\cdot\sum_{t=1}^{T}U_tF'_t\right)C\end{aligned}$$

Using (3.59),

$$(\text{1st-term}) = O\left(\frac{1}{\sqrt{NT}}\right) \ , \quad (\text{2nd-term}) = O\left(\frac{1}{\sqrt{NT}}\right) \tag{3.67}$$

As for the 3rd term,

$$(\text{3rd-term}) = \alpha_i' \left(\frac{1}{T} \cdot \sum_{t=1}^{T} U_t U_t'\right) = O\left(\frac{1}{N}\right) \tag{3.68}$$

which is obtained from (3.56). Regarding the 4th term, it is easy to show

$$E\left[\frac{1}{T} \cdot \sum_{t=1}^{T} F_t V_{it}'\right] = 0 \ , \quad Var\left[\frac{1}{T} \cdot \sum_{t=1}^{T} F_t V_{it}'\right] = \frac{k\sigma_v^2 I_m}{T}$$

because F_t and V_{it} are independent. Therefore, let us suppose

$$(\text{4th-term}) = \beta' \left[\frac{1}{T} \cdot \sum_{t=1}^{T} V_{it} F_t'\right] C = O\left(\frac{1}{\sqrt{T}}\right) \tag{3.69}$$

From (3.65), the 5th term is:

$$(\text{5th-term}) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right) \tag{3.70}$$

From (3.62),

$$(\text{6th-term}) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right) \tag{3.71}$$

From (3.56),

$$(\text{7th-term}) = O\left(\frac{1}{N}\right) \tag{3.72}$$

Therefore, from (3.66)-(3.72),

$$\psi_1^{(a)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{T}}\right)$$

$\psi_1^{(b)}$ (in Lemma 10)

$$\psi_1^{(b)} = \frac{1}{T} \cdot \sum_{t=1}^{T} (C' F_t U_t' + U_t F_t' C') \ + \ \frac{1}{T} \cdot \sum_{t=1}^{T} U_t U_t'$$

From (3.59), the 1st term of RHS is

$$\text{(1st-term)} = O\left(\frac{1}{\sqrt{NT}}\right)$$

Also, from (3.56),

$$\text{(2nd-term)} = O\left(\frac{1}{N}\right)$$

Therefore,

$$\psi_1^{(b)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right)$$

<u>$\psi_2^{(a)}$ **(in Lemma 10)**</u>

$$\begin{aligned}\psi_2^{(a)} &= \left(\frac{1}{T}\cdot\sum_{t=1}^{T} V_{it}F_t'C\right) + \left(\frac{1}{T}\cdot\sum_{t=1}^{T} \Gamma_i' F_t U_t'\right) \\ &\quad + \left(\frac{1}{T}\cdot\sum_{t=1}^{T} V_{it}U_t'\right)\end{aligned}$$

As for the 1st term of RHS,

$$\text{(1st-term)} = O\left(\frac{1}{\sqrt{NT}}\right)$$

because V_{it} and F_t are independent. Also, from (3.59),

$$\text{(2nd-term)} = O\left(\frac{1}{\sqrt{NT}}\right)$$

From (3.65),

$$\text{(3rd-term)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right)$$

Therefore,

$$\psi_2^{(a)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right)$$

$\underline{\Delta_\Sigma^{(p)} \textbf{ (in Lemma 12)}}$

$$
\begin{aligned}
\Delta_\Sigma^{(p)} \quad = \quad & \frac{1}{NT} \cdot \sum_{i,t} \left[\widetilde{X}_{it}^{(p)} (\psi_2 Z_t)' + (\psi_2 Z_t) \widetilde{X}_{it}^{(p)\prime} \right] \\
& - \frac{1}{NT} \cdot \sum_{i,t} \left[V_{it} U_t' (C'C)^{-1} C' \Gamma_i + \Gamma_i' C (C'C)^{-1} U_t' V_{it}' \right] \\
& + \frac{1}{NT} \cdot \sum_{i,t} \Gamma_i' C (C'C)^{-1} U_t U_t' (C'C)^{-1} C' \Gamma_i' \\
& - \frac{1}{NT} \cdot \sum_{i,t} \psi_2 Z_t Z_t' \psi_2'
\end{aligned}
$$

As for the 1st term of RHS, Lemma 11 gives

$$(\text{1st-term}) = 0$$

From (3.65),

$$(\text{2nd-term}) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right)$$

From (3.56),

$$(\text{3rd-term}) = O\left(\frac{1}{N}\right)$$

As for the last term,

$$(\text{4th-term}) = O(\psi_2^2) = O\left(\frac{1}{T}\right) + O\left(\frac{1}{N^2}\right)$$

From all the above,

$$\Delta_\Sigma^{(p)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{T}\right)$$

$\underline{\delta_1^{(a)}}$ **(in Lemma 14)**

$$\begin{aligned}
\delta_1^{(a)} &= \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}(\alpha_i'C'+\beta'\Gamma_i')F_tU_t' \;+\; \frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\alpha_i'U_tF_t'C \\
&\quad + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\alpha_i'U_tU_t' + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\beta'V_{it}F_t'C \\
&\quad + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\beta'V_{it}U_t' \;+\; \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\eta_{it}U_t' \\
&= \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}(\alpha_i'C'+\beta'\Gamma_i')F_tU_t' \;+\; \frac{1}{NT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\alpha_i'U_tF_t'C \qquad (3.73)\\
&\quad + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\alpha_i'U_tU_t' + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\beta'V_{it}F_t'C \\
&\quad + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\beta'V_{it}U_t' + \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\varepsilon_{it}U_t' \\
&\quad - \frac{1}{N_jT}\cdot\sum_{t=1}^{T}\sum_{i\in G_j}\alpha_i'U_tU_t'
\end{aligned}$$

As for the 1st and 2nd terms of RHS,

$$\begin{aligned}
\text{(1st-term)} &= \frac{1}{N_j}\cdot\sum_{i\in G_j}(\alpha_i'C'+\beta'\Gamma_i')\left[\frac{1}{T}\cdot\sum_{t=1}F_tU_t'\right] \\
\text{(2nd-term)} &= \frac{1}{T}\cdot\sum_{i\in G_j}\alpha_i'\left(\frac{1}{T}\cdot\sum_{t=1}^{T}U_tF_t'\right)C
\end{aligned}$$

Using (3.59),

$$\text{(1st-term)} = O\left(\frac{1}{\sqrt{NT}}\right)\ ,\quad \text{(2nd-term)} = O\left(\frac{1}{\sqrt{NT}}\right) \qquad (3.74)$$

As for the 3rd term,

$$\text{(3rd-term)} = \frac{1}{N_j}\cdot\sum_{i\in G_j}\alpha_i'\left(\frac{1}{T}\cdot\sum_{t=1}^{T}U_tU_t'\right) = O\left(\frac{1}{N}\right) \qquad (3.75)$$

which is obtained from (3.56). Regarding the 4th term, it is easy to show

$$E\left[\frac{1}{N_jT}\cdot\sum_{i\in G_j}\sum_{t=1}^{T}F_tV_{it}'\right] = 0\ ,\quad Var\left[\frac{1}{N_jT}\cdot\sum_{i\in G_j}\sum_{t=1}^{T}F_tV_{it}'\right] = \frac{k\sigma_v^2 I_m}{N_jT}$$

because F_t and V_{it} are independent. Therefore, let us suppose

$$(\text{4th-term}) = \beta' \left[\frac{1}{N_j T} \cdot \sum_{i \in G_j} \sum_{t=1}^{T} V_{it} F_t' \right] C = O\left(\frac{1}{\sqrt{NT}} \right) \tag{3.76}$$

Furthermore, by using (3.65), (3.54) and (3.56), 5th, 6th and 7th terms are obtained as

$$\begin{aligned} &(\text{5th-term}) = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right) \\ &(\text{6th-term}) = O\left(\frac{1}{N}\right) \quad , \quad (\text{7th-term}) = O\left(\frac{1}{N}\right) \end{aligned} \tag{3.77}$$

Therefore, from (3.73)-(3.77),

$$\delta_1^{(a)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right)$$

<u>$\delta_1^{(b)}$ **(in Lemma 14)**</u>

$$\delta_1^{(b)} = \frac{1}{T} \cdot \sum_{t=1}^{T} \left(C' F_t U_t' + U_t F_t' C' \right) \; + \; \frac{1}{T} \cdot \sum_{t=1}^{T} U_t U_t'$$

From (3.59), the 1st term of RHS is

$$(\text{1st-term}) = O\left(\frac{1}{\sqrt{NT}}\right)$$

Also, from (3.56),

$$(\text{2nd-term}) = O\left(\frac{1}{N}\right)$$

Therefore,

$$\delta_1^{(b)} = O\left(\frac{1}{N}\right) + O\left(\frac{1}{\sqrt{NT}}\right)$$

<u>$\delta_2^{(a)}$ **(in Lemma 14)**</u>

$$\begin{aligned} \delta_2^{(a)} \quad = \quad & \left(\frac{1}{N_j T} \cdot \sum_{t=1}^{T} \sum_{i \in G_j} V_{it} F_t' C \right) + \left(\frac{1}{N_j T} \cdot \sum_{t=1}^{T} \sum_{i \in G_j} \Gamma_i' F_t U_t' \right) \\ & + \left(\frac{1}{N_j T} \cdot \sum_{t=1}^{T} \sum_{i \in G_j} V_{it} U_t' \right) \end{aligned}$$

As for the 1st term of RHS, it is easy to show

$$E\left[\frac{1}{N_jT}\cdot\sum_{i\in G_j}\sum_{t=1}^{T}F_tV_{it}'\right]=0\ ,\quad Var\left[\frac{1}{N_jT}\cdot\sum_{i\in G_j}\sum_{t=1}^{T}F_tV_{it}'\right]=\frac{k\sigma_v^2I_m}{N_jT}$$

because F_t and V_{it} are independent. Therefore, let us suppose

$$(\text{1st-term})=O\left(\frac{1}{\sqrt{NT}}\right)$$

because V_{it} and F_t are independent. Also, from (3.59),

$$(\text{2nd-term})=O\left(\frac{1}{\sqrt{NT}}\right)$$

From (3.65),

$$(\text{3rd-term})=O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{NT}}\right)$$

Therefore,

$$\delta_2^{(a)}=O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{NT}}\right)$$

<u>$\Delta_\Sigma^{(X)}$ **(in Lemma 16)**</u> In (3.39),

$$\begin{aligned}\Delta_\Sigma^{(X)} \quad = \quad & \frac{1}{NT}\cdot\sum_{i,t}\left[\widetilde{X}_{it}(\delta_2Z_t)'+(\delta_2Z_t)\widetilde{X}_{it}'\right]\\
&+\frac{1}{NT}\cdot\sum_{i,t}\left[(\Gamma_i'-\overline{\Gamma}')F_tV_{it}'+V_{it}F_t'(\Gamma_i-\overline{\Gamma})\right]\\
&-\frac{1}{NT}\cdot\sum_{i,t}\Big[(\Gamma_i'-\overline{\Gamma}')F_t\overline{U}_t'(C'C)^{-1}C'\overline{\Gamma}\\
&\qquad\qquad+\overline{\Gamma}'C(C'C)^{-1}\overline{U}_tF_t'(\Gamma_i-\overline{\Gamma})\Big]\\
&-\frac{1}{NT}\cdot\sum_{i,t}\left[V_{it}\overline{U}_t'(C'C)^{-1}C'\overline{\Gamma}+\overline{\Gamma}'C(C'C)^{-1}\overline{U}_tV_{it}'\right]\\
&+\frac{1}{NT}\cdot\sum_{i,t}\overline{\Gamma}'C(C'C)^{-1}\overline{U}_t\overline{U}_t'(C'C)^{-1}C'\overline{\Gamma}\\
&-\frac{1}{NT}\cdot\sum_{i,t}\delta_2Z_tZ_t'\delta_2'\end{aligned} \tag{3.78}$$

As for the 1st term of RHS,

$$\frac{1}{NT}\cdot\sum_{i,t}\widetilde{X}_{it}(\delta_2Z_t)'=0 \tag{3.79}$$

which is obtained from Lemma 15. Since V_{it} is independent of $\{F_t, \Gamma_i\}$,

$$\frac{1}{NT} \cdot \sum_{i,t} (\Gamma_i' - \overline{\Gamma}') F_t V_{it}' = O\left(\frac{1}{\sqrt{NT}}\right) \tag{3.80}$$

In the 3rd term of RHS,

$$\begin{aligned}
&\frac{1}{NT} \cdot \sum_{i,t} (\Gamma_i' - \overline{\Gamma}') F_t \overline{U}_t' (C'C)^{-1} C' \overline{\Gamma} \\
&= \left(\frac{1}{N} \cdot \sum_{i=1}^{N} (\Gamma_i' - \overline{\Gamma})\right) \left(\frac{1}{T} \cdot \sum_{t=1}^{T} F_t \overline{U}_t'\right) (C'C)^{-1} C' \overline{\Gamma} \\
&= (0_{1\times k}) \left(\frac{1}{T} \cdot \sum_{t=1}^{T} F_t \overline{U}_t'\right) (C'C)^{-1} C' \overline{\Gamma} \\
&= 0
\end{aligned} \tag{3.81}$$

In the 4th term,

$$\begin{aligned}
\frac{1}{NT} \cdot \sum_{i,t} V_{it} \overline{U}_t' (C'C)^{-1} C' \overline{\Gamma} &= \left(\frac{1}{T} \cdot \sum_{t=1}^{T} V_t \overline{U}_t'\right) (C'C)^{-1} C' \overline{\Gamma} \\
&= O\left(\frac{1}{N}\right) \qquad \text{(from (3.55))}
\end{aligned} \tag{3.82}$$

In the 5th term,

$$\begin{aligned}
&\frac{1}{NT} \cdot \sum_{i,t} \overline{\Gamma}' C (C'C)^{-1} \overline{U}_t \overline{U}_t' (C'C)^{-1} \overline{\Gamma} \\
&\qquad = \overline{\Gamma}' C (C'C)^{-1} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} \overline{U}_t \overline{U}_t'\right] (C'C)^{-1} C' \overline{\Gamma} \\
&\qquad = O\left(\frac{1}{N}\right) \qquad \text{(from (3.56))}
\end{aligned} \tag{3.83}$$

In the 6th term,

$$\begin{aligned}
\frac{1}{NT} \cdot \sum_{i,t} \delta_2 Z_t Z_t' \delta_2' &= \delta_2 \left(\frac{1}{T} \cdot \sum_{t=1}^{T} Z_t Z_t'\right) \delta_2' \\
&= O\left(\delta_2^2\right) = O\left(\frac{1}{N^2}\right) + O\left(\frac{1}{NT}\right)
\end{aligned} \tag{3.84}$$

From (3.78)-(3.84),

$$\Delta_{\Sigma}^{(X)} = O\left(\frac{1}{\sqrt{NT}}\right) + O\left(\frac{1}{N}\right)$$

References for Chapter 3

Ando, T. and Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31:163–191.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77:1229–1279.

Belloni, A., Chernozhukov, V., and Wang, L. (2011). Square-root Lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika*, 98:791–806.

Bohomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83:1147–1184.

Bonhomme, S., Lamadon, T., and Manresa, E. (2017). Discretizing unobserved heterogeneity. IFS Working Papers W17/03, Institute for Fiscal Studies.

Gu, J. and Volgushev, S. (2018). Panel data quantile regression with grouped fixed effects. Technical Report arXiv:1801.05041, ArXiv.

Hocking, T. D., Joulin, A., Bach, F., and Vert, J.-P. (2011). Clusterpath: an algorithm for clustering using convex fusion penalties. In *ICML'11 Proceedings of the International Conference on Machine Learning*, pages 745–752.

Lin, C.-C. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1:1–14.

Lu, X. and Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190:148–175.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83:1543–1579.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33:158–195.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74:967–1012.

Radchenko, P. and Mukherjee, G. (2017). Convex clustering via ℓ_1 fusion penalization. *Journal of the Royal Statistical Society Series B*, 79:1527–1546.

Sun, Y. X. (2005). Estimation and inference in panel structure models. Economics Working Paper Series 9t5tf1231k, University of California at San Diego.

Tan, K. M. and Witten, D. (2015). Statistical properties of convex clustering. *Electronic Journal of Statistics*, 9:2324–2347.

Zhu, C., Xu, H., Leng, C., and Yan, S. (2014). Convex optimization procedure for clustering: theoretical revisit. In *Advances in Neural Information Proceedings Systems*, pages 1619–1627.